

Introdução à acessibilidade urbana

um guia prático em R

Rafael H. M. Pereira, Daniel Herszenhut

2022-10-17T00:00:00+00:00

Table of contents

Apresentação	5
Organização do livro	6
Código reproduzível	6
Licença de uso	6
Como citar o livro	7
Agradecimentos	7
 I SEÇÃO 1: Introdução à acessibilidade urbana	 8
1 O que é acessibilidade?	10
1.1 Definição de acessibilidade urbana	10
1.2 Diferença entre microacessibilidade e acessibilidade urbana	11
1.3 Por que a acessibilidade urbana importa?	12
1.4 Diferença entre acessibilidade e mobilidade	12
 2 Medidas de acessibilidade	 15
2.1 Medidas baseadas em lugares	15
2.1.1 Custo mínimo de viagem	16
2.1.2 Medida de oportunidades cumulativas	16
2.1.3 Medidas gravitacionais	17
2.1.4 Medidas de acessibilidade com competição: <i>floating catchment area</i> . . .	18
2.2 Medidas baseadas em pessoas	19
 II SEÇÃO 2: Calculando acessibilidade	 20
3 Calculando acessibilidade urbana em R	22
3.1 Cálculo da matriz de tempo de viagem	22
3.1.1 Instalação do r5r	22
3.1.2 Dados necessários	23
3.1.3 Calculando a matriz de tempo de viagens	24
3.2 Cálculo de acessibilidade	27
3.3 Cálculo de acessibilidade com o r5r	30
3.4 Análises de acessibilidade	31

III	SEÇÃO 3: Dados de transporte público	37
4	Dados GTFS	39
4.1	Estrutura de arquivos GTFS	39
4.1.1	agency.txt	41
4.1.2	stops.txt	42
4.1.3	routes.txt	43
4.1.4	trips.txt	44
4.1.5	calendar.txt	45
4.1.6	shapes.txt	46
4.1.7	stop_times.txt	47
4.1.8	frequencies.txt	48
4.2	Onde encontrar GTFS de cidades brasileiras	50
5	Manipulação e visualização de dados GTFS	51
5.1	Leitura e manipulação básica de arquivos GTFS	51
5.2	Cálculo de velocidade das linhas	53
5.3	Combinando e filtrando <i>feeds</i>	55
5.4	Mapeamento do <i>headway</i> das linhas	69
6	Comparando a acessibilidade entre dois cenários de transporte	76
6.1	Arquivos GTFS usados na análise	77
6.2	Cálculo das matrizes de tempo de viagem	78
6.3	Cálculo da acessibilidade nos cenários antes e depois	83
6.4	Análise dos níveis de acessibilidade antes e depois	86
6.5	Desigualdade de acessibilidade	92
V	SEÇÃO 5: Dados do Projeto AOP	97
	Quais dados estão disponíveis?	98
7	Dados de população e socioeconômicos	102
7.1	Download dos dados	102
7.2	Mapa de população total	103
7.3	Mapa de população por cor	104
7.4	Mapa de população por renda	106
8	Dados de distribuição espacial de oportunidades	108
8.1	Download dos dados	108
8.2	Mapa de empregos	109
8.3	Mapa de escolas	110
8.4	Mapa de estabelecimentos de saúde	111
8.5	Mapa de CRAS	112

9 Estimativas e mapas de acessibilidade	114
9.1 Download dos dados	114
9.2 Mapa do tempo para acessar o hospital mais próximo	120
9.3 Mapa da quantidade de oportunidades acessíveis	122
9.4 Desigualdades de acesso a oportunidades	123
Referências bibliográficas	129

Apresentação

O objetivo deste livro é apresentar uma introdução sobre os conceitos e habilidades práticas necessárias para fazer estudos e avaliações de impacto sobre acessibilidade urbana. O livro começa dando uma visão geral sobre o conceito e indicadores de acessibilidade. Em seguida, ensina como analisar dados espaciais e de redes de transporte para se calcular estimativas de acessibilidade por diferentes modos de transporte e como visualizar esses resultados em mapas e gráficos. O livro também busca familiarizar o leitor com dados de redes de transporte público em formato GTFS e ensina como manipular e visualizar esse tipo de dado. Num dos seus principais capítulos, o livro ensina ainda como os dados e metodologia apresentados nos capítulos anteriores podem ser utilizados para avaliar o impacto de projetos de transporte sobre as condições de acesso a oportunidades da população. Por fim, o livro ensina como baixar, analisar e visualizar estimativas de acessibilidade já calculadas para cidades brasileiras e disponibilizadas pela equipe do Projeto Acesso a Oportunidades (AOP) do Ipea.

O tema da acessibilidade tem recebido cada vez mais atenção de agências de transporte, instituições de financiamento, tomadores de decisão e pesquisadores na área de planejamento urbano e de transportes ([papa2015accessibility?](#); [boisjoly2017how?](#)). Com isso, existe um crescente número de artigos científicos ([miller2018accessibility?](#); [van2021transport?](#)) e livros ([levine2019mobility?](#); [levinson2020transport?](#)) que trazem rico material para discussão e aprofundamento sobre questões relacionadas à acessibilidade urbana. No entanto, não existem na literatura livros ou artigos que sirvam simultaneamente de material didático introdutório sobre o tema e de manual prático de metodologia e análise de dados. A ausência desse tipo de material explica, ao menos em parte, porque diversas agências de transporte e analistas ainda enfrentam desafios em incorporar questões e indicadores de acessibilidade no seu dia-a-dia de planejamento e pesquisa ([silva2017accessibility?](#); [buttner2021accessibility?](#)).

Este livro foi escrito pensando nas necessidades de trabalho de gestores públicos, analistas, alunos e pesquisadores de planejamento e transporte urbano. Por isso, o livro tem caráter prático. Todo material do livro é apresentado com exemplos reproduzíveis e dados abertos utilizando-se programação em R. O objetivo é equipar seus leitores com os conceitos fundamentais e as ferramentas de análise e processamento de dados necessárias para realizar análises de acessibilidade urbana e conduzir avaliações do impacto de projetos e políticas de transporte a partir de dados e programas de código abertos.

Organização do livro

Este livro está dividido em 5 seções:

1. Seção 1: A primeira seção apresenta o conceito de acessibilidade urbana, esclarece a diferença entre acessibilidade e mobilidade urbana, e apresenta os principais indicadores usados para medir a acessibilidade urbana.
2. Seção 2: A segunda seção ensina como calcular estimativas de acessibilidade urbana em R usando os pacotes `{r5r}` e `{accessibility}` a partir de dados públicos abertos.
3. Seção 3: A terceira seção apresenta o que é a especificação GTFS de dados de transporte público, e mostra como trabalhar e analisar dados de GTFS usando o pacote `{gtfstools}`.
4. Seção 4: A quarta seção mostra como o conhecimento dos capítulos anteriores pode ser utilizado para se avaliar o impacto de políticas urbanas sobre condições de acessibilidade urbana usando R.
5. Seção 5: Por fim, a quinta seção mostra como fazer download e analisar os dados do projeto Acesso a Oportunidades (AOP) para analisar a distribuição espacial e socioeconômica das condições de acesso a oportunidades nas cidades brasileiras.

Código reproduzível

Este livro traz diversos exemplos reproduzíveis de análise de dados em R. A maneira mais fácil de você rodar todo o código do livro no seu navegador (*web browser*) é utilizando o [Binder](#). Clicando neste link, você será direcionado para uma sessão do RStudio dentro do seu navegador onde você poderá abrir e rodar os scripts de todos os capítulos do livro.

i Este livro foi escrito e publicado com o sistema de publicação [Quarto](#). Todo o código utilizado em seu preparo e na sua publicação pode ser encontrado [neste repositório](#).

Se você quiser rodar o conteúdo do livro localmente na sua máquina, você pode clonar o repositório do livro e ter o RStudio instalado. Após abrir o projeto `intro_access_book.Rproj`, basta rodar a função:

```
quarto::quarto_preview().
```

Licença de uso

É permitida a reprodução e a exibição para uso educacional ou informativo, desde que respeitado o crédito ao autor original e citada a fonte. Permitida a inclusão da obra em Repositórios ou Portais de Acesso Aberto, desde que fique claro para os usuários os termos

de uso da obra e quem é o detentor dos direitos autorais, o Instituto de Pesquisa Econômica Aplicada (Ipea). Proibido o uso comercial ou com finalidades lucrativas em qualquer hipótese. Proibida a criação de obras derivadas. Para imagens estáticas e em movimento (vídeos e audiovisuais), ATENÇÃO: os direitos de imagem foram cedidos apenas para a obra original, formato de distribuição e repositório. Esta licença está baseada em estudos sobre a Lei Brasileira de Direitos Autorais (Lei 9.610/1998).

Como citar o livro

Caso você use o material deste livro, pedimos por gentileza que você cite o livro utilizando a seguinte referência:

- Pereira, Rafael H. M. & Herszenhut, Daniel. (2022) Introdução à acessibilidade urbana um guia prático em R. Instituto de Pesquisa Econômica Aplicada.

Agradecimentos

Este livro foi elaborado pelo Instituto de Pesquisa Econômica Aplicada (Ipea) no âmbito da parceria entre Ipea e a Secretaria de Mobilidade e Desenvolvimento Regional e Urbano (SMDRU) do Ministério do Desenvolvimento Regional (MDR).

Part I

SEÇÃO 1: Introdução à acessibilidade urbana

Objetivo: O objetivo desta seção é (1) apresentar o conceito de acessibilidade urbana, esclarecendo a diferença entre acessibilidade e mobilidade, e (2) apresentar uma visão geral dos principais indicadores usados para medir acessibilidade.

Quanto a postos de trabalho uma pessoa em um determinado local consegue alcançar em até uma hora de viagem usando transporte público? Quanto tempo essa pessoa leva para chegar até o posto de saúde ou a escola mais próxima de sua casa? As respostas a essas perguntas dependem diretamente das políticas locais de transporte e de desenvolvimento urbano. Essas políticas determinam os níveis de acessibilidade urbana em cada cidade - isto é, a facilidade com que cada indivíduo consegue acessar oportunidades de emprego, serviços de saúde e educação, atividades culturais e de lazer, entre outros tipos de atividade. A acessibilidade, portanto, é simultaneamente um resultado da conectividade e performance dos sistemas de transporte, da organização espacial das cidades em termos da distribuição da população, atividades econômicas e serviços públicos por seu território. Assim, a promoção da acessibilidade tem papel-chave para o desenvolvimento econômico, e para promoção de cidades mais sustentáveis e inclusivas. Ao longo desta seção vamos entender com mais detalhes o conceito de acessibilidade, por que ele é importante para entendermos o funcionamento das cidades e quais os indicadores mais comumente utilizados para medir a acessibilidade urbana.

1 O que é acessibilidade?

1.1 Definição de acessibilidade urbana

Acessibilidade é a facilidade com que as pessoas conseguem alcançar lugares e oportunidades – ou, inversamente, uma característica de lugares e oportunidades em termos do quão facilmente eles podem ser alcançados pela população (geurs2004accessibility?; neutens2010equity?).

As condições de acessibilidade são influenciadas tanto pela co-distribuição espacial da população e de atividades econômicas e serviços públicos quanto pela configuração e desempenho da rede de transportes. Nesse sentido, a acessibilidade urbana tem papel fundamental na capacidade das pessoas de se deslocarem para acessar oportunidades, como empregos, escolas, etc.

Os níveis de acessibilidade urbana são estabelecidos, portanto, por três componentes distintos:

- **Infraestrutura:** A facilidade de acessar atividades depende da infraestrutura e dos serviços de transporte existentes. Isso inclui, por exemplo, a capilaridade da rede de transporte público, a conectividade da rede viária, a existência de corredores de transporte de alta capacidade, como trens e metrô, etc. Aqui, tanto a eficiência quanto a conectividade espacial e temporal da rede de transportes são de extrema importância.
- **Uso do solo:** O quão fácil atividades podem ser acessadas também depende da co-distribuição espacial de pessoas e atividades, como escolas, serviços de saúde, áreas de lazer, etc. Esse componente diz respeito à proximidade geográfica entre pessoas e oportunidades: quanto mais longe, mais difícil é o acesso às atividades.
- **Pessoas:** Por fim, não podemos esquecer que a facilidade de acessar atividades também é afetada pelas características individuais de cada pessoa. Fatores como dificuldades motoras e cognitivas, idade, gênero, cor e renda, por exemplo, podem influenciar de maneira importante a capacidade das pessoas se locomoverem, de utilizarem determinados modos de transporte e de circularem pela cidade sem medo de algum tipo de violência ou discriminação.

A consideração desses fatores traz importantes informações para análises sobre equidade e inclusão social. No entanto, a sua influência sobre as condições de acessibilidade das pessoas costuma ser melhor captada em análises qualitativas. Por conta de dificuldades operacionais

e computacionais, essa dimensão da acessibilidade costuma receber pouca atenção de estudos de larga escala de avaliação de impacto de projetos de transporte. No capítulo Chapter 2, nós ponderamos algumas vantagens e desvantagens de indicadores que conseguem captar essas três dimensões da acessibilidade urbana.

1.2 Diferença entre microacessibilidade e acessibilidade urbana

Para fins de esclarecimento de conceitos, é importante distinguir entre o que nós chamamos de acessibilidade urbana e o que em geral se entende por acessibilidade, em português.

O termo *acessibilidade* é comumente utilizado para se referir a questões de padrões e normas de design universal e de construção e planejamento para inclusão de pessoas com diferentes graus de dificuldades motoras e cognitivas. Isso é o que normalmente entende-se por **microacessibilidade**, pois abrange questões de acesso a serviços e atividades na escala micro - ou seja, como detalhes do planejamento de espaços públicos e privados, e a construção de veículos e equipamentos o, por exemplo, afetam a capacidade de cada indivíduo de conseguir acessar lugares, serviços, produtos, etc.

A *acessibilidade urbana*, por sua vez, pode ser entendida como **macroacessibilidade**, pois trata de questões de acesso de uma maneira mais ampla. Ou seja, quando falamos de acessibilidade urbana nos concentramos em como questões estruturais de planejamento e desenvolvimento urbano, como a disposição de corredores de transportes e a distribuição espacial de pessoas e atividades, afetam a capacidade das pessoas de acessar oportunidades. Ela trata de como a capacidade de acessar atividades considerando tanto as capacidades das pessoas utilizarem tecnologias de transporte quanto a distribuição espacial de atividades vis-à-vis a cobertura e conectividade da rede de transporte.

Notadamente, a microacessibilidade e a macroacessibilidade são elementos complementares de uma noção mais ampla de acessibilidade. Condições de microacessibilidade, por exemplo, afetam diretamente a capacidade de pessoas embarcarem e utilizarem diferentes modos de transporte, de se locomoverem com segurança sobre calçadas, de atravessarem ruas, etc. De pouco adianta um indivíduo morar em uma região atendida por uma grande oferta de modos de transporte se a pessoa é cadeirante, por exemplo, e a rede de transporte não é adaptada para pessoas com dificuldade de locomoção. Neste livro focaremos apenas em análises de acessibilidade urbana e muitas vezes utilizaremos o termo acessibilidade como um sinônimo de macroacessibilidade por uma questão de fluência do texto. É importante reconhecer, no entanto, que a macroacessibilidade de forma isolada não é suficiente para descrever as condições de acessibilidade de um indivíduo em sua totalidade, e que para tal precisaríamos olhar também para a microacessibilidade.

1.3 Por que a acessibilidade urbana importa?

O conceito de acessibilidade é central em estudos de transporte por diferentes razões. Primeiramente, este conceito articula de maneira explícita como políticas de transporte e políticas de desenvolvimento e uso do solo urbano interagem e impactam a capacidade das pessoas de se deslocarem nas cidades. Por sua vez, o acesso a oportunidades e atividades, como postos de trabalho e serviços de educação e de saúde, por exemplo, tem papel fundamental para a satisfação das necessidades individuais e sociais e para promoção da inclusão social (**pereira_equity_2021?**; **luz2022understanding?**). A promoção da acessibilidade também é uma condição necessária, embora não suficiente, para a expansão da liberdade de escolha das pessoas (**church2000transport?**; **lucas2016transport?**; **vanwee2022accessibility?**). O conceito de acessibilidade, portanto, serve como fio condutor para relacionar investimentos e políticas de transportes e uso do solo a questões relacionadas à exclusão social e qualidade de vida de indivíduos, como o potencial de satisfação de necessidades básicas e a liberdade.

Adicionalmente, a ideia de acessibilidade traz à tona a dimensão espacial de um problema central de justiça social, que é a desigualdade de oportunidades. Ela ajuda, portanto, a explicitamente incorporar a noção de espaço no desenho de políticas destinadas a enfrentar essas injustiças (**farrington2005rural?**; **pereira2017distributive?**). Assim, a noção de acessibilidade se torna peça fundamental para se pensar a equidade de políticas públicas, e de se avaliar os aspectos distributivos de quais grupos sociais e quais bairros se beneficiam dessas políticas.

Como vimos anteriormente, os níveis de acesso a oportunidades numa cidade são resultado conjunto da capacidade de cada pessoa utilizar tecnologias de transporte e da integração entre a distribuição geográfica de atividades, a distribuição espacial da população na cidade e a conectividade espacial e temporal da rede de transportes (**miller2018accessibility?**; **paez2012measuring?**). O planejamento com foco na acessibilidade, desta forma, passa por um planejamento integrado entre o uso do solo e o sistema de transportes, buscando criar maior proximidade entre pessoas e atividades e reduzindo a dependência de modos de transporte motorizados (**banister2011trilogy?**). Planejar as cidades e sistemas de transporte visando melhorar a acessibilidade, portanto, é essencial para o desenvolvimento de cidades inclusivas e sustentáveis.

1.4 Diferença entre acessibilidade e mobilidade

A diferença entre os conceitos de acessibilidade e mobilidade é frequentemente ignorada, mesmo por pesquisadores e planejadores que lidam com esses temas diariamente. Afinal de contas, existe uma grande intersecção temática entre a acessibilidade e o que se entende de por “mobilidade urbana” enquanto área de pesquisa e política pública: uma área que lida com os deslocamentos das pessoas, que está relacionada ao planejamento de sistemas de transporte coletivo e individual, com o planejamento de redes ciclovárias e de pedestres, etc. Neste contexto, não é

raro, por exemplo, ouvir que um determinado grupo populacional “tem menos mobilidade” do que outro, quando na verdade o que se quer dizer é que esse grupo apresenta piores condições de acessibilidade. Qual é, então, a diferença entre acessibilidade e mobilidade?

Na literatura científica de planejamento urbano e de transportes, o conceito de mobilidade diz respeito aos padrões de viagens que as pessoas efetivamente fazem no seu dia a dia – por exemplo, quantas viagens foram feitas, quais modos de transporte foram usados, qual a distância média das viagens, quanto tempo se gasta no deslocamento casa-trabalho, etc. Informações de mobilidade são tradicionalmente captadas por meio de pesquisas domiciliares origem-destino. Recentemente, com o surgimento de novas tecnologias digitais, dados de GPS de telefones celulares, de cartões de bilhetagem eletrônica, de radares e semáforos urbanos, entre outros, vêm também sendo usados com a finalidade de descrever os deslocamentos diários da população (**anda2017transport?**; **kandt2021smart?**). Dados e análises de mobilidade trazem informações sobre as condições diárias de uso do sistema de transportes e sobre os padrões de viagens de indivíduos de diferentes grupos socioeconômicos, o que permite captar aspectos importantes do desempenho econômico e ambiental das cidades e do bem-estar da população.

O conceito de acessibilidade, por sua vez, está intrinsecamente relacionado ao *potencial* que as pessoas têm de alcançar atividades e oportunidades. Enquanto uma análise de mobilidade foca, por exemplo, no que as pessoas levam no seu deslocamento diário de sua casa até o seu trabalho, uma análise de acessibilidade tenta identificar, digamos, se as pessoas conseguiriam alcançar seus empregos em um tempo de viagem tido como razoável ou qual a quantidade de empregos potencialmente alcançáveis dentro de um determinado custo de viagem. A acessibilidade trata do quão fácil/factível é alcançar um local, a mobilidade trata de como esse local foi alcançado. Níveis de acessibilidade são, portanto, medidas potenciais, ao passo que os dados de mobilidade descrevem padrões reais, realizados.

Tradicionalmente, o planejamento urbano e de transportes tem como foco a mobilidade, (**banister2011trilogy?**; **vasconcellos2018urban?**; **levinson2020transport?**). Ainda hoje, o foco na mobilidade motiva políticas que priorizam a circulação de automóveis e visam aumentar a velocidade e a fluidez de trânsito para reduzir congestionamentos e, conseqüentemente, reduzir tempos de deslocamento [**levine2019mobility**]. Essas políticas, no entanto, tendem a dar um enfoque quantitativo na mobilidade: aumentar o número de viagens, aumentar a velocidade média, diminuir o tempo de congestionamento, etc. Neste contexto, portanto, a mobilidade é vista como um fim em si mesma, e “melhorar” a mobilidade dependeria de políticas como soluções puramente técnicas que deveriam “otimizar” os aspectos quantitativos mencionados anteriormente.

A mobilidade, no entanto, não pode ser encarada como uma finalidade em si. As pessoas raramente se deslocam pelo simples prazer de se deslocar. Ao contrário, na grande maioria das vezes as pessoas se deslocam como um meio para acessar as atividades no destino da viagem. Neste sentido, tem-se observado um crescente consenso entre pesquisadores e agências de transporte de que o objetivo de uma política de transportes é melhorar o acesso da população (**pereira2017distributive?**; **martens2012justice?**; **bertolini2005sustainable?**).

Se o que as pessoas querem é acessar atividades, então é possível pensar em formas de planejamento que facilitem seu acesso a tais atividades sem necessariamente promover o aumento da motorização e da velocidade no trânsito, que implicam no aumento de externalidades econômicas, ambientais, de saúde pública, entre outras.

Neste caso, há uma mudança de paradigma no planejamento urbano e de transportes: a busca por padrões de viagem mais sustentáveis implica na mudança do foco da mobilidade para a acessibilidade (**banister2008sustainable?**). Políticas de aumento de velocidade nas vias e de expansão de faixas, por exemplo, podem ser substituídas por políticas de promoção de maior *mix* de uso do solo e de aproximação de pessoas e atividades, promovendo maior integração entre planejamento de transporte e uso do solo. Quando o foco da política muda da mobilidade para a acessibilidade urbana, abre-se um leque maior de possíveis instrumentos e ações de políticas públicas para promover um desenvolvimento urbano mais integrado e calcado na promoção da sustentabilidade e da inclusão social (**banister2011trilogy?**; **levine2019mobility?**).

2 Medidas de acessibilidade

A mudança de paradigma no planejamento urbano e de transportes para a promoção da acessibilidade traz com si diversos desafios. Entre eles está a necessidade de criar métodos que busquem medir as condições de acessibilidade nas cidades. A busca por estimativas de acessibilidade facilmente comunicáveis, metodologicamente robustas e pouco intensivas computacionalmente levou ao desenvolvimento de um grande número de diferentes medidas (**paez2012measuring?**). Essas medidas podem ser divididas em dois grandes grupos: medidas baseadas em lugares e medidas baseadas em pessoas (**dijst2002opportunities?**).

2.1 Medidas baseadas em lugares

Medidas baseadas em lugares medem a acessibilidade como uma característica de um determinado local. Por simplificação, esses indicadores assumem que todas as pessoas que se encontram em um mesmo local têm as mesmas condições de acesso às atividades distribuídas pela cidade. Ou seja, se uma análise de acessibilidade utiliza uma medida baseada em lugares para calcular a acessibilidade e divide a área de estudo em uma grade quadriculada, cada célula dessa grade (um quadrado) terá a ela associado um nível de acessibilidade, que é posteriormente atribuído a todas as pessoas que residem dentro daquela célula. Essas medidas são sensíveis a fatores relacionados à distribuição espacial de atividades e à configuração e desempenho da rede de transporte, mas não levam em consideração as características individuais das pessoas.

As medidas desse tipo são as mais amplamente utilizadas por agências de transporte e pesquisadores (**papa2015accessibility?**). Isto porque elas exigem menor quantidade de dados e tendem a ser consideravelmente mais fáceis de serem calculadas e interpretadas do que suas contrapartes baseadas em pessoas. Por este motivo, os exemplos e estudos de caso apresentados neste nos próximos capítulos deste livro focam somente neste grupo de medidas.

Medidas de acessibilidade baseadas em lugares associam a cada deslocamento um custo, usualmente expresso em termos de viagem (**el-geneidy2016cost?**; **venter2016assessing?**) - ou seja, se um local pode ser alcançado a partir de outro em meia hora, o custo para realizar essa viagem é de 30 minutos. Nada impede, no entanto, que outros tipos de custo, como a distância da viagem, seu custo monetário e a percepção de conforto dos usuários, por exemplo, sejam considerados (**arbex2020estimating?**; **herszenhut2022impact?**). Nós apresentamos abaixo uma rápida descrição de algumas das medidas baseadas em lugares mais

frequentemente utilizadas na literatura científica e na prática de agências de transporte. O termo “custo” é utilizado aqui de maneira ampla, e pode se referir a qualquer tipo de unidade quantitativa utilizada para quantificar a impedância de uma viagem, seja ela tempo de viagem, custo monetário ou demais alternativas.

2.1.1 Custo mínimo de viagem

Uma das medidas de acessibilidade mais simples, a medida de custo mínimo de viagem indica qual é o menor custo de necessário para alcançar a oportunidade mais próxima a partir de uma determinada origem. Ela permite captar, por exemplo, qual é o tempo de viagem até o posto de saúde mais próximo a partir de cada quarteirão de uma cidade. O indicador é calculado com a seguinte fórmula:

$$A_i = \min(c_{i1}, c_{i2}, \dots, c_{ij}, \dots, c_{i(n-1)}, c_{in}) \iff O_j \geq 1$$

em que A_i é a acessibilidade na origem i , c_{ij} é o custo de deslocamento entre a origem i e o destino j , n é o número total de destinos na área de estudo e O_j é o número de oportunidades no destino j .

Vantagens e desvantagens: Esta medida tem as vantagens de ser fácil de se calcular e exigir poucos dados, além de ser fácil de comunicar. Duas desvantagens, no entanto, são que ela não capta a quantidade de oportunidades acessíveis e nem aspectos de competição na demanda pela oportunidade. Por exemplo, uma pessoa pode morar muito perto de um hospital, mas essa proximidade pode não garantir um bom acesso aos serviços de saúde se esse for o único hospital da região e estiver sujeito a picos de demanda que sobrecarregam os serviços além de suas capacidades.

2.1.2 Medida de oportunidades cumulativas

A medida de oportunidades cumulativas mede a quantidade de oportunidades que pode ser alcançada dentro de um determinado limite de custo de viagem. Por exemplo, este indicador pode ser utilizado para medir a quantidade de empregos acessíveis por transporte público em até 60 minutos ou a quantidade de escolas acessíveis em até 30 minutos de viagem a pé. A medida é calculada com a seguinte fórmula:

$$A_i = \sum_{j=1}^n O_j \times f(c_{ij})$$

$$f(c_{ij}) = \begin{cases} 1 & \text{se } c_{ij} \leq C \\ 0 & \text{caso contrário} \end{cases}$$

em que A_i é a acessibilidade na origem i , O_j é o número de oportunidades no destino j , n é o número total de destinos na área de estudo, $f(c_{ij})$ é uma função binária que assume os valores 0 ou 1, a depender do custo de deslocamento c_{ij} entre a origem i e o destino j , e C é o limite de custo de deslocamento estabelecido.

Vantagens e desvantagens: A medida de oportunidades cumulativas também é fácil de se calcular, exige poucos dados, e é fácil de comunicar. Isso contribui para que este indicador seja um dos mais utilizados por agências de transporte e de financiamento em análises de acessibilidade (**papa2015accessibility?**; **boisjoly2017how?**). Entre as suas desvantagens estão o fato de que este indicador não considera a influência da competição sobre oportunidades e exige a escolha de um único ponto de corte como custo máximo de viagem. Além disso, esta medida assume que todas as oportunidades são igualmente desejáveis e alcançáveis pelas pessoas, desde que os custos de viagem estejam dentro do limite pré-estabelecido - ou seja, uma oportunidade a 10 minutos de viagem é igualmente acessível a outra que esteja a 40 minutos de viagem, por exemplo, se o custo limite for de 60 minutos.

2.1.3 Medidas gravitacionais

Mais do que um tipo de medida específica, podemos entender as medidas gravitacionais como uma família de medidas. Assim como no caso da medida de oportunidades cumulativas, esta família de métricas considera a soma das oportunidades que podem ser alcançadas a partir de um determinado local. A contagem de oportunidade, no entanto, é gradualmente descontada à medida que o custo da viagem aumenta. Assim, oportunidades mais fáceis de acessar têm uma importância maior, e o peso de cada oportunidade na soma total diminui com a sua dificuldade de acesso em relação à origem.

O ritmo de decaimento desse peso em função do custo da viagem é ditado por uma função de impedância, que pode ser definida de diferentes formas. Por exemplo, a **função de decaimento linear** considera que o peso de cada oportunidade diminui de maneira constante ao longo do trajeto até um determinado custo limite, onde ele passa a ser zero. Já a **função de decaimento exponencial negativa** considera que o peso de cada oportunidade é dividido por um fator que cresce exponencialmente, fazendo com que ele diminua rapidamente a custos de viagem baixos e se aproxime de 0 a custos altos. As fórmulas abaixo apresentam a formulação genérica de uma medida gravitacional e as funções de decaimento linear e exponencial negativa, mencionadas acima.

$$A_i = \sum_{j=1}^n O_j \times f(c_{ij})$$

$$f_{lin}(c_{ij}) = \begin{cases} 1 - c_{ij}/C & \text{se } c_{ij} \leq C \\ 0 & \text{caso contrário} \end{cases}$$

$$f_{exp}(c_{ij}) = e^{-\beta c_{ij}}$$

em que A_i é a acessibilidade na origem i , O_j é o número de oportunidades no destino j , n é o número total de destinos na área de estudo, $f(c_{ij})$ é uma função de decaimento cujo resultado varia com o custo de deslocamento c_{ij} entre a origem i e o destino j , $f_{lin}(c_{ij})$ é a função de decaimento linear, C é o limite de custo de deslocamento estabelecido, $f_{exp}(c_{ij})$ é a função de decaimento exponencial negativa e β é um parâmetro que dita a velocidade de decaimento.

Um sem-número de diferentes funções de decaimento podem ser utilizadas no cálculo de medidas gravitacionais. A medida de oportunidades cumulativas, por exemplo, pode ser pensada simplesmente como um caso especial de medida gravitacional em que o peso de cada oportunidade é ditado por uma função binária, em vez de decair gradualmente. (levinson2020transport?) apresentam uma lista de funções de decaimento utilizadas por agências de transportes e pesquisadores em análises que envolvem medidas gravitacionais.

Vantagens e desvantagens: A principal vantagem de medidas gravitacionais é que o desconto do peso das oportunidades pelo custo da viagem reflete, de alguma maneira, a forma como as pessoas costumam perceber o acesso a oportunidades: serviços e atividades costumam ser mais atrativas quanto mais próximos eles estiverem, tudo mais constante. Este indicador, no entanto, tem ao menos duas desvantagens. A primeira delas é que os níveis de acessibilidade estimados são de difícil interpretação, pela forma como a contagem de oportunidades é descontada pelo custo de viagem. Além disso, para que esses níveis sejam mais representativos do comportamento de viagem das pessoas, o ritmo de decaimento da função de impedância (o parâmetro β da função exponencial negativa, por exemplo) precisa ser calibrado. Por isso, esta medida requer a disponibilidade de dados de comportamento de viagens ou outros dados que possam ser usados no processo de calibração, disponíveis, por exemplo, a partir de pesquisas tipo origem-destino, dados de telefone celular etc.

2.1.4 Medidas de acessibilidade com competição: *floating catchment area*

Em muitos casos, o acesso a oportunidades é afetado não apenas por questões de proximidade geográfica e de custos de transporte, mas também pela competição de diferentes pessoas por uma mesma oportunidade. Isso é muito comum, por exemplo, em casos de acesso a serviços de saúde, escolas e empregos. Uma vaga de emprego só pode ser ocupada por uma pessoa de cada vez, e o mesmo vale para um leito de UTI ou uma vaga em uma escola.

Existe uma gama de medidas de acessibilidade que buscam incorporar essa competição pelas oportunidades na estimativa dos níveis de acessibilidade. Algumas das medidas de competição mais amplamente utilizadas são as do tipo *floating catchment area* (FCA, em tradução livre área de influência flutuante). A título de exemplo, esses indicadores tentam levar em consideração como uma mesma pessoa pode potencialmente acessar vários leitos de UTI e, simultaneamente, como cada leito de UTI pode ser acessado por diversas pessoas. Assim, o acesso de uma

pessoa ao serviço de leito de UTI é influenciado simultaneamente tanto por questões de custos de transporte como também pela disponibilidade de leitos de UTI considerando a potencial concorrência na demanda por aqueles leitos.

Dentro da família de medidas de FCA, a mais comumente utilizada é a *2-Step Floating Catchment Area* (2SFCA), proposta originalmente por Luo and Wang (2003). Uma limitação da 2SFCA é que ela contabiliza que uma mesma pessoa pode demandar várias oportunidades ao mesmo tempo e que, analogamente, um mesmo serviço pode ser utilizado por várias pessoas ao mesmo tempo. Esses fenômenos são conhecidos como problemas de inflação de demanda e de oferta, respectivamente, e podem gerar estimativas de acessibilidade enviesadas ou pouco precisas (Paez, Higgins, and Vivona 2019). Para lidar com esses problemas, Paez, Higgins, and Vivona (2019) propuseram a *Balanced Floating Catchment Area* (BFCA), uma das mais novas da família de medidas de FCA.

Vantagens e desvantagens: Diferentes medidas de FCA têm diferentes vantagens e desvantagens, em maior ou menor grau. No entanto, de maneira geral, a principal vantagem de medidas desta família é a sua capacidade de incorporar aspectos de competição em estimativas de acessibilidade. A principal desvantagem, por outro lado, é a difícil interpretação e comunicação dos seus resultados.

2.2 Medidas baseadas em pessoas

Medidas de acessibilidade baseadas em pessoas são sensíveis não apenas à distribuição espacial de atividades e à configuração e desempenho da rede de transporte. Indicadores desta categoria também levam em consideração como que características pessoais de cada indivíduo (como sexo, idade, deficiência física, etc.) e até questões como participação em atividades e compromissos pessoais podem afetar a facilidade de acesso de uma pessoa a determinadas atividades. Esta categoria inclui, por exemplo, medidas de acessibilidade baseadas em utilidade (**miller2018accessibility?**), indicadores baseadas em atividades (**dong2006moving?**) e medidas de espaço-tempo (**neutens2012analysis?**).

Vantagens e desvantagens: Embora indicadores desta categoria sejam mais sofisticados, eles costumam demandar grandes quantidades de dados, incluindo registros de diários de viagem, pesquisas domiciliares tipo origem-destino, etc. Por isso, o cálculo dessas medidas é computacionalmente mais intensivo e complexo, o que faz com que elas sejam menos utilizadas do que as medidas baseadas em lugares (**miller2018accessibility?**).

Part II

SEÇÃO 2: Calculando acessibilidade

Objetivo: O objetivo desta seção é mostrar como calcular estimativas de acessibilidade urbana em R usando os pacotes `{r5r}` e `{accessibility}`.

O cálculo dos níveis de acessibilidade em um determinado local compreende duas etapas principais: primeiro, nós precisamos calcular uma matriz de custo, geralmente o tempo de viagem, entre as origens e os destinos de uma determinada cidade ou região considerando um modo de transporte; feito isso, calculamos a acessibilidade em cada ponto de origem considerando os custos de transporte entre cada origem e o número de oportunidades em cada destino. Nesta seção nós aprenderemos mais sobre essas duas etapas usando R, quais dados são necessários para executá-las e quais as vantagens e desvantagens dos diferentes métodos que podem ser usados para isso.

3 Calculando acessibilidade urbana em R

3.1 Cálculo da matriz de tempo de viagem

A primeira etapa necessária para estimar os níveis de acessibilidade de uma área urbana é calcular a matriz de custo de viagem entre as diversas origens e destinos que a compõem. Como comentado anteriormente, na literatura científica e na prática do planejamento de sistemas de transporte esse custo é mais frequentemente representado pelo tempo de viagem que separa dois pontos ([el-geneidy2016cost?](#); [venter2016assessing?](#)), embora trabalhos recentes tenham considerado também outros fatores, como o dinheiro necessário para realizar uma viagem e o nível de conforto da viagem entre um ponto e outro ([arbex2020estimating?](#); [herszenhut2022impact?](#)). Pela prevalência deste tipo de matriz na literatura e na prática, porém, iremos nos focar em matrizes de tempo de viagem.

Atualmente, a forma mais fácil e rápida de gerar uma matriz de tempo de viagem em R é utilizando o pacote `{r5r}` ([pereira2021r5r?](#)), desenvolvido pela equipe do Projeto Acesso a Oportunidades, do Ipea. O pacote utiliza, por trás dos panos, o *software* de roteamento de transporte multimodal R5, desenvolvido pela Conveyal¹.

3.1.1 Instalação do r5r

A instalação do `{r5r}` funciona como a instalação de qualquer pacote de R (todos os exemplos de código a seguir devem ser rodados em uma sessão de R).

```
install.packages("r5r")
```

Além do R, o pacote `{r5r}` requer também a instalação do Java 11². Use o comando abaixo para checar a versão do Java instalada em seu computador.

```
cat(processx::run("java", args = "-version")$stdout)
```

¹Disponível em <https://www.pasfor.com.br/>.

²Mais detalhes sobre o uso e a sintaxe do `data.table` podem ser lidos em <https://rdatatable.gitlab.io/data.table/index.html>.

Como podemos ver, a versão instalada no livro é compatível com o `{r5r}`. Caso a versão instalada na sua máquina não seja (resultado do código mencionando a versão 12 ou 1.8.0, por exemplo), por favor atualize o Java para a versão 11.

3.1.2 Dados necessários

O uso do pacote `{r5r}` requer diferentes tipos de dados. A lista abaixo descreve cada um deles, comenta sobre sua obrigatoriedade e apresenta algumas fontes onde esses dados podem ser obtidos:

- **Rede viária** (obrigatório): um arquivo com a rede viária e de infraestrutura de pedestres do *OpenStreetMap*, em formato `.pbf`. Pode ser baixado com:
 - [osmextract](#), pacote de R;
 - [Geofabrik](#), *website*;
 - [HOT Export Tool](#), *website*;
 - [BBBike Extract Service](#), *website*.
- **Rede de transporte público** (opcional): um arquivo GTFS descrevendo a rede de transporte público da área de estudo. Caso ausente, deslocamentos por transporte público não podem ser roteados. Pode ser baixado com:
 - [tidytransit](#), pacote de R;
 - [Transitland](#), *website*;
 - No capítulo Chapter 4 deste livro (Table 4.9) nós indicamos também onde baixar os dados de GTFS de algumas cidades brasileiras que compartilham seus dados publicamente.
- **Topografia** (opcional): um arquivo de dados *raster*, com o modelo digital de elevação em formato `.tif`. Utilizado caso deseje-se levar em consideração os efeitos da topografia do local sobre os tempos de caminhada e de viagens de bicicleta. Pode ser baixado com:
 - [elevatr](#), pacote de R;
 - Nasa's [SRTMGL1](#), *website*.

Esses dados devem ser salvos em uma mesma pasta que, preferencialmente, não contenha nenhum outro arquivo. Como veremos adiante, o `{r5r}` combina todos os dados salvos nesta pasta para criar uma rede de transporte multimodal que é utilizada para roteamento e cálculo das matrizes de tempo de viagem. Note que é possível ter mais de um arquivo GTFS na mesma pasta: nesse caso, o `{r5r}` considera as redes de transporte públicos de todos os *feeds* de forma conjunta. No entanto, a pasta deve conter um único arquivo de rede viária. Assumindo que os scripts de R estarão em uma pasta chamada R, uma possível organização dos arquivos segue o esquema abaixo:

```

/tmp/RtmpRCldcx/projeto_acessibilidade
+-- R
|   +-- script1.R
|   \-- script2.R
\-- r5
    +-- rede_transporte_publico.zip
    +-- rede_viaria.osm.pbf
    \-- topografia.tif

```

Para ilustrar as funcionalidades do {r5r}, nós vamos usar uma pequena amostra de dados da cidade de Porto Alegre, Brasil. Esses dados estão disponíveis dentro do próprio pacote {r5r}, e seu endereço pode ser acessado com o comando `system.file("extdata/poa", package = "r5r")`:

```

/home/runner/work/intro_access_book/intro_access_book/renv/library/R-4.2/x86_64-pc-linux-gnu
+-- poa.zip
+-- poa_elevation.tif
+-- poa_hexgrid.csv
+-- poa_osm.pbf
\-- poa_points_of_interest.csv

```

Esta pasta possui quatro arquivos que vamos usar agora:

- A rede viária do *OpenStreetMap*: `poa_osm.pbf`;
- Dois *feeds* de GTFS das redes de ônibus e de trens: `poa_eptc.zip` e `poa_trensurb.zip`;
- O modelo digital de elevação da cidade: `poa_elevation.tif`;
- O arquivo `poa_hexgrid.csv`, com as coordenadas geográficas dos centróides de uma grade hexagonal regular que cobre toda a área de estudo e com informações sobre o tamanho da população residente e o número de oportunidades (empregos, escolas e hospitais) em cada hexágono. Esses pontos serão utilizados como as origens e destinos no cálculo da matriz de tempo de viagem.

3.1.3 Calculando a matriz de tempo de viagens

Antes de calcular a matriz de tempo de viagem, precisamos aumentar a memória disponível para o Java. Isto é necessário porque, por padrão, o R aloca apenas 512MB de memória para processos do Java, o que frequentemente não é suficiente para o cálculo de grandes matrizes com o {r5r}. Para aumentar a memória disponível para 2GB, por exemplo, usamos o seguinte comando no início do *script*, antes mesmo de carregar as bibliotecas de R necessárias:

```
options(java.parameters = "-Xmx2G")
```


Feito isso, agora podemos carregar as bibliotecas que vamos utilizar neste capítulo.

```
library(r5r)
library(accessibility)
library(sf)
library(data.table)
library(ggplot2)
library(aopdata)
```

O cálculo de uma matriz de tempo de viagem de uma área de estudo é feito em dois passos. O primeiro é gerar a rede de transporte multimodal que será utilizada no roteamento. Para isso, nós utilizamos a função `setup_r5()`, que baixa o *software* de roteamento R5 e o utiliza para criar a rede. Esta função recebe como *input* o caminho da pasta onde os dados necessários para o roteamento estão armazenados. Como resultado, a função salva na pasta alguns arquivos necessários para o roteamento e retorna uma conexão com o R5, neste exemplo chamada de `r5r_core`, que é utilizada no cálculo da matriz de tempo de viagem.

```
path <- system.file("extdata/poa", package = "r5r")

r5r_core <- setup_r5(path, use_elevation = TRUE, verbose = FALSE)

fs::dir_tree(path)
```

```
/home/runner/work/_temp/renv/cache/v5/R-4.2/x86_64-pc-linux-gnu/r5r/0.7.1/61db001154cd833466
+-- network.dat
+-- poa.zip
+-- poa_elevation.tif
+-- poa_hexgrid.csv
+-- poa_osm.pbf
+-- poa_osm.pbf.mapdb
+-- poa_osm.pbf.mapdb.p
\-- poa_points_of_interest.csv
```

O passo final é usar a função `travel_time_matrix()` para o cálculo da matriz de tempo de viagem. Como *inputs* básicos, a função recebe a conexão com o R5 criada acima, pontos de origem e destino em formato de `data.frame` com as colunas `id`, `lon` e `lat`, o modo de transporte a ser utilizado e o horário de partida, o tempo máximo de caminhada permitido da origem até o embarque no transporte público e do desembarque até o destino, e o tempo máximo de viagem a ser considerado.

```

points <- fread(file.path(path, "poa_hexgrid.csv"))

ttm <- travel_time_matrix(
  r5r_core,
  origins = points,
  destinations = points,
  mode = c("WALK", "TRANSIT"),
  departure_datetime = as.POSIXct(
    "13-05-2019 14:00:00",
    format = "%d-%m-%Y %H:%M:%S"
  ),
  max_walk_dist = 800,
  max_trip_duration = 120,
  verbose = FALSE,
  progress = FALSE
)

head(ttm)

```

	fromId	toId	travel_time
1:	89a901291abffff	89a901291abffff	1
2:	89a901291abffff	89a901295b7ffff	41
3:	89a901291abffff	89a9012809bffff	43
4:	89a901291abffff	89a901285cfffff	35
5:	89a901291abffff	89a90e934d7ffff	71
6:	89a901291abffff	89a90129b47ffff	58

Na prática, o que a função `travel_time_matrix()` faz é encontrar a rota mais rápida partindo de cada ponto de origem para todos os possíveis pontos de destino, considerando o modo de viagem, o horário de partida e os demais parâmetros passados pelo usuário. Para isso, o `{r5r}` considera tempos de viagem de porta-a-porta: no caso de uma viagem por transporte público, por exemplo, o tempo total de viagem inclui a) o tempo de caminhada até a parada de transporte público; b) o tempo de espera pelo veículo na parada; c) o tempo de deslocamento dentro do veículo; e d) o tempo de viagem a pé da parada de desembarque até o destino. Em casos em que mais de uma rota de transporte público é utilizada, o `{r5r}` também contabiliza o tempo gasto nas conexões, considerando a caminhada entre paradas e o tempo de espera pelo próximo veículo.

Diversos outros *inputs* podem ser passados para o cálculo da matriz, como a velocidade de caminhada e o número máximo de pernas de transporte público permitido, entre outros. Para mais informações sobre cada um dos parâmetros, por favor consulte a documentação da

função disponível no pacote em uma sessão de R (com os comandos `?travel_time_matrix()` e `help("travel_time_matrix")`) ou no site do `{r5r}`, [aqui](#).

i A função `travel_time_matrix()` utiliza uma extensão do algoritmo de roteamento RAPTOR ([conway2017evidencebased?](#)), o que torna o R5 extremamente rápido. A depender da quantidade de pares de origem-destino, o `{r5r}` pode ser entre 6 e 200 vezes mais rápido do que outros *softwares* de roteamento multimodal no cálculo de uma matriz de tempo de viagem ([higgins2022calculating?](#)).

3.2 Cálculo de acessibilidade

Após calculada a matriz de tempo de viagem entre as origens e os destinos da área de estudo, nós precisamos utilizá-la para calcular os níveis de acessibilidade do local. Para isso, nós utilizaremos o pacote `{accessibility}`³, também desenvolvido pela equipe do Projeto Acesso a Oportunidades, Ipea, que contém diversas funções para o cálculo de vários indicadores de acessibilidade. Como *input* básico, todas as funções requerem uma matriz de custo pré-calculada (no nosso caso, a matriz de tempo de viagem calculada na seção anterior) e dados de uso do solo, como, por exemplo, o número de determinados tipos de oportunidades em cada célula da grade que cobre a área de estudo.

Medida de oportunidades cumulativas

A função `cumulative_cutoff()` é utilizada para calcular o número de oportunidades que pode ser alcançado em um determinado limite de custo de viagem. No exemplo abaixo, nós calculamos o número de escolas que podem ser alcançados em até 30 minutos de viagem a partir de cada origem presente em nossa matriz de tempo de viagem.

```
# renomear nome das colunas
setnames(ttm, c("fromId", "toId"), c("from_id", "to_id"))

cumulative_access <- cumulative_cutoff(
  travel_matrix = ttm,
  land_use_data = points,
  opportunity = "schools",
  travel_cost = "travel_time",
  cutoff = 30
)

head(cumulative_access)
```

³Disponível em <https://github.com/ipeaGIT/accessibility>.

	id	schools
1:	89a901291abffff	19
2:	89a9012a3cffff	0
3:	89a901295b7ffff	15
4:	89a901284a3ffff	0
5:	89a9012809bffff	20
6:	89a901285cffff	82

Custo mínimo de viagem

A função `cost_to_closest()`, por sua vez, calcula o custo mínimo de viagem necessário para alcançar um determinado número de oportunidades. O código abaixo, por exemplo, calcula o tempo de viagem mínimo para alcançar o hospital mais próximo a partir de cada origem.

```
min_cost <- cost_to_closest(
  travel_matrix = ttm,
  land_use_data = points,
  opportunity = "schools",
  travel_cost = "travel_time"
)

head(min_cost)
```

	id	travel_time
1:	89a9012124ffff	0
2:	89a9012126bffff	16
3:	89a9012127bffff	11
4:	89a90128003ffff	7
5:	89a90128007ffff	20
6:	89a9012800bffff	0

Medidas gravitacionais

A função `gravity()` calcula medidas gravitacionais de acessibilidade - ou seja, aquelas nas quais o peso de cada oportunidade diminui gradualmente com o aumento do custo de viagem. Existem, no entanto, uma gama de diferentes tipos de funções de decaimento que podem ser utilizadas, como funções de decaimento exponenciais negativas, de potências inversas, entre outras. Por isso, esta função recebe um *input* adicional: a função de decaimento a ser utilizada no cálculo. O exemplo abaixo apresenta o cálculo de acessibilidade a estabelecimentos de educação usando uma medida gravitacional exponencial negativa com parâmetro de decaimento igual a 0.2.

```
negative_exp_access <- gravity(
  travel_matrix = ttm,
  land_use_data = points,
  opportunity = "schools",
  travel_cost = "travel_time",
  decay_function = decay_exponential(0.2)
)

head(negative_exp_access)
```

```
      id  schools
1: 89a901291abffff 0.3589287
2: 89a9012a3cfffff 0.0000000
3: 89a901295b7ffff 0.5206127
4: 89a901284a3ffff 0.0000000
5: 89a9012809bffff 0.4476908
6: 89a901285cfffff 1.9573827
```

Medidas com competição

Por fim, a função `floating_catchment_area()` calcula níveis de acessibilidade levando em consideração a competição por oportunidades usando diferentes indicadores do tipo *floating catchment area*. Como diversos métodos de FCA podem ser utilizados, a função requer que o método desejado seja explicitamente assinalado. E, assim como a função de acessibilidade gravitacional, a função de decaimento utilizada também deve ser definida pelo usuário. O código a seguir mostra um exemplo de cálculo de acessibilidade a oportunidades de emprego usando o método BFCA (Paez, Higgins, and Vivona 2019), levando em consideração os efeitos de competição entre a população como um todo e uma função de decaimento exponencial com parâmetro de decaimento igual a 0.05.

```
bfca_access <- floating_catchment_area(
  travel_matrix = ttm,
  land_use_data = points,
  opportunity = "schools",
  travel_cost = "travel_time",
  demand = "population",
  method = "bfca",
  decay_function = decay_exponential(0.05)
)

head(bfca_access)
```

	id	schools
1:	89a901291abffff	0.0002574442
2:	89a9012a3cffff	0.0000000000
3:	89a901295b7ffff	0.0002091911
4:	89a901284a3ffff	0.0000000000
5:	89a9012809bffff	0.0002276971
6:	89a901285cffff	0.0003928648

As funções apresentadas nesta seção também podem receber outros *inputs* não explicitamente mencionados aqui. Para mais informações sobre cada um dos parâmetros, por favor consulte a documentação do pacote `{accessibility}` em seu [site](#).

3.3 Cálculo de acessibilidade com o `r5r`

Ao longo das duas seções anteriores, nós mostramos como calcular níveis de acessibilidade passo-a-passo. Para fins didáticos, é importante entender que o cálculo de estimativas de acessibilidade tem como primeiro passo a geração de uma matriz de custos de viagens que posteriormente é utilizada para estimar níveis de acessibilidade. No entanto, o `{r5r}` inclui também uma função chamada `accessibility()` que faz todo esse processamento de uma vez só.

De forma parecida com a função de cálculo de matriz de tempo de viagem, a função `accessibility()` recebe como *inputs* uma conexão com o R5, as origens, os destinos, os modos de transporte e o tempo de partida, entre outros argumentos. Adicionalmente, devem ser listadas também quais oportunidades devem ser consideradas e a função de decaimento que deve ser utilizada, bem como o valor do limite de custo e do parâmetro de decaimento. O exemplo abaixo mostra uma aplicação desta função.

```
r5r_access <- accessibility(
  r5r_core,
  origins = points,
  destinations = points,
  opportunities_colname = "schools",
  decay_function = "step",
  cutoffs = 30,
  mode = c("WALK", "TRANSIT"),
  departure_datetime = as.POSIXct(
    "13-05-2019 14:00:00",
    format = "%d-%m-%Y %H:%M:%S"
  ),
  max_walk_dist = 800,
```

```

    max_trip_duration = 120,
    verbose = FALSE,
    progress = FALSE
)

head(r5r_access)

```

	from_id	percentile	cutoff	accessibility
1:	89a901291abffff	50	30	17
2:	89a9012a3cffff	50	30	0
3:	89a901295b7ffff	50	30	13
4:	89a901284a3ffff	50	30	0
5:	89a9012809bffff	50	30	17
6:	89a901285cffff	50	30	78

Como podemos observar, o resultado desta função são os mesmos níveis de acessibilidade já calculados anteriormente, no exemplo da medida de oportunidades cumulativas. A informação “intermediária” do tempo de viagem entre origens e destinos, consequentemente, não fica disponível ao usuário com o uso desta função. Ainda assim, este fluxo de trabalho pode ser uma boa alternativa para pessoas que estejam interessadas unicamente nos níveis de acessibilidade, não dependendo do tempo de viagem em suas análises.

3.4 Análises de acessibilidade

Calculados os níveis de acessibilidade, procedemos então para sua análise. Existe uma grande variedade de análises que podem ser feitas usando esses dados: por exemplo, em diagnósticos das condições de acessibilidade urbana de diferentes bairros, pesquisas sobre desigualdades de acesso a oportunidades entre diferentes grupos sociais, análises sobre exclusão social e *accessibility poverty* (insuficiência de acessibilidade, em tradução livre), etc. Nesta seção, no entanto, nos restringiremos a apresentar duas análises relativamente simples e de fácil comunicação: a distribuição espacial da acessibilidade e sua distribuição entre diferentes grupos de renda.

Distribuição espacial de acessibilidade urbana

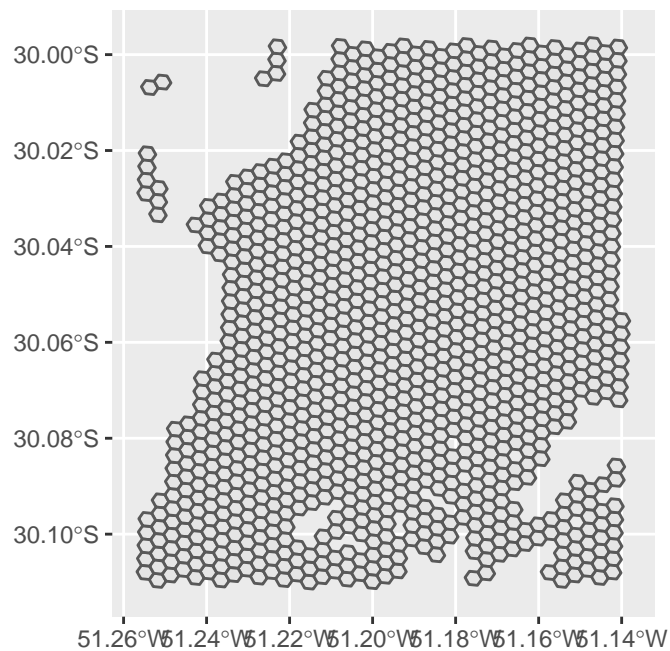
Para compreendermos a distribuição espacial da acessibilidade urbana de uma determinada cidade ou região, primeiro precisamos obter as informações espaciais dos pontos que foram utilizados como origens e destinos no cálculo da matriz. Os pontos que nós usamos nos exemplos anteriores, por exemplo, correspondem a células de uma grade hexagonal baseadas no índice H3, desenvolvido pela Uber ([brodsky2018h3?](#)). O código e o mapa a seguir mostram a distribuição desses hexágonos na nossa área de estudo. Aqui, a gente faz o download da

grade espacial disponível no pacote {aopdata}, que será apresentado com mais detalhes na seção ?@sec-s5_dados_aop;

```
# download grade espacial
poa_grid <- aopdata::read_grid(city="Porto Alegre")

poa_grid <- poa_grid[poa_grid$id_hex %in% points$id, ]

# cria mapa
ggplot(poa_grid) + geom_sf()
```



Para visualizarmos os nossos dados de acessibilidade espacialmente, portanto, nós precisamos unir a nossa tabela de estimativas de acessibilidade (vamos usar os níveis calculados com a medida de oportunidades cumulativas) com a tabela que contém os dados espaciais da grade, usando as colunas de identificação dos hexágonos como colunas-chave. A operação de *left join* usada para isso, bem como o seu resultado em formato de imagem, são apresentados a seguir.

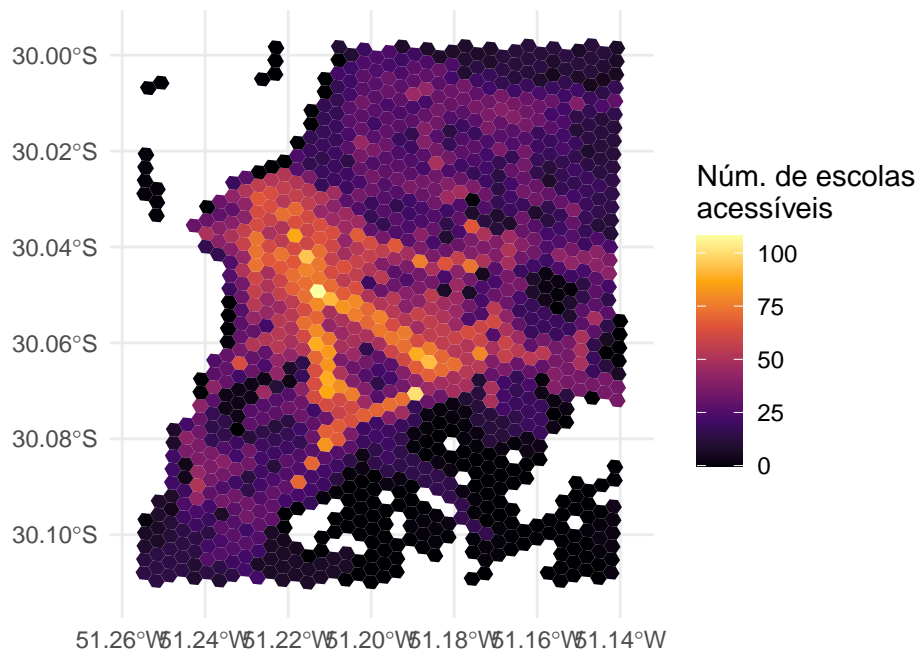
```
# converte objeto para data.table
setDT(poa_grid)
```



```
# merge para trazer info do número de escolas
poa_grid[cumulative_access, on = c(id_hex = "id"), schools_access := i.schools]

# converte de volta para objeto espacial sf
poa_grid <- st_sf(poa_grid)

# plotar mapa
ggplot(poa_grid) +
  geom_sf(aes(fill = schools_access), color = NA) +
  scale_fill_viridis_c(option = "inferno") +
  labs(fill='Núm. de escolas\acessíveis') +
  theme_minimal()
```



Como podemos ver, os níveis de acessibilidade tendem a se concentrar de forma mais acentuada no centro da cidade, onde existe maior concentração de empregos, e próximos aos grandes corredores de transportes da cidade. Por terem fácil acesso a modos de alta capacidade e velocidade, pessoas que moram mais perto desses corredores tendem a acessar locais distantes relativamente rápido. Em contraste, pessoas que moram mais longe desses corredores, dependem de modos de menor frequência e velocidade operacional (como os ônibus municipais, por exemplo) e precisam gastar mais tempo para alcançar os corredores de média e alta capacidade. Como consequência, os níveis de acessibilidade de pessoas que moram afastadas do centro e de corredores de alta capacidade tendem a ser menores.

Distribuição socioeconômica de acessibilidade urbana

O mapa acima, embora revelador quanto aos locais em que estão dispostas as maiores concentrações de acessibilidade, nada diz sobre quais são os grupos socioeconômicos que possuem os maiores potenciais de acesso a oportunidades na região. Para isso, nós precisamos cruzar informações demográficas e econômicas das pessoas que moram em cada um dos nossos pontos de origem com os dados de acessibilidade previamente calculados.

No exemplo abaixo, nós juntamos aos dados de acessibilidade a informação do decil de renda de cada uma das origens, considerando a renda média de cada uma das pessoas que as habitam. Assim, nós conseguimos identificar se um hexágono é de baixa, média ou alta renda. Os dados socioeconômicos que utilizamos no exemplo vêm do censo demográfico brasileiro de 2010, e foram agregados na grade espacial de hexágonos por ([pereira2022usosolo?](#)). Aqui, nós acessamos os dados no R usando o pacote `{aopdata}`, que será apresentado em mais detalhes na seção `?@sec-s5_dados_aop`.

```
poa_population <- aopdata::read_population("Porto Alegre", showProgress = FALSE)

setDT(poa_grid)
poa_grid[
  poa_population,
  on = "id_hex",
  `:=`(
    pop_count = i.P001,
    income_decile = i.R003
  )
]

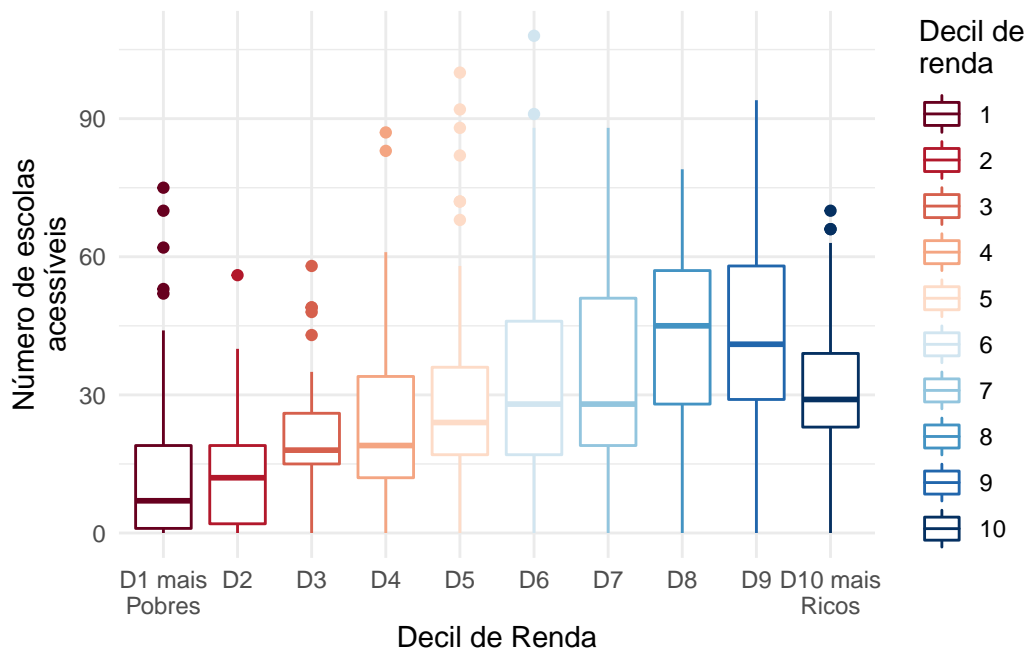
poa_grid[, .(id_hex, schools_access, pop_count, income_decile)] |> head()
```

	id_hex	schools_access	pop_count	income_decile
1:	89a9012124fffff	14	733	9
2:	89a9012126bffff	13	355	9
3:	89a9012127bffff	14	996	10
4:	89a90128003ffff	34	1742	4
5:	89a90128007ffff	15	477	5
6:	89a9012800bffff	34	501	4

Tendo a informação do decil de renda em que cada hexágono se encontra, nós podemos calcular a distribuição da acessibilidade da população dentro de cada um desses níveis de renda. Para isso, nós precisamos ponderar o nível de acessibilidade de cada origem pela quantidade de pessoas que residem ali - daí o porquê de termos também trazido a informação da contagem populacional em cada hexágono. Fazendo a ponderação, nós obtemos a distribuição

da acessibilidade das pessoas localizadas em origens de um determinado decil de renda. Caso não ponderássemos, no entanto, nós teríamos a distribuição de acessibilidade não das pessoas localizadas em cada hexágono, mas dos hexágonos em si. Como em nossa análise nós nos importamos com as pessoas, e não com as unidades espaciais em que elas estão agregadas, precisamos fazer a ponderação. Nós podemos visualizar a distribuição de acessibilidade de cada decil usando um *boxplot*, como feito a seguir.

```
ggplot(poa_grid[!is.na(income_decile)]) +
  geom_boxplot(
    aes(
      x = as.factor(income_decile),
      y = schools_access,
      color = as.factor(income_decile),
      weight = pop_count
    )
  ) +
  labs(color="Decil de\renda", x="Decil de Renda", y="Número de escolas\nacessíveis") +
  scale_color_brewer(palette = 'RdBu') +
  scale_x_discrete(labels=c("D1 mais\nPobres", paste0('D', 2:9), "D10 mais\nRicos")) +
  theme_minimal()
```



O gráfico é muito claro em seu conteúdo: pessoas de mais baixa renda tendem a ter níveis de

acessibilidade consideravelmente menores do que as de alta renda. Isso é um padrão comum em praticamente todas as cidades brasileiras (R. H. M. Pereira et al. 2019). Isto ocorre, em larga medida, devido à localização espacial das comunidades de baixa e alta renda no território: os mais ricos costumam morar em áreas mais valorizadas, próximas das grandes concentrações de empregos (e oportunidades de educação, saúde, lazer, etc., também) e com maior oferta de transporte público de média e alta capacidade. Os mais pobres, por sua vez, tendem a morar em locais mais afastados, onde o valor da terra é menor. Consequentemente, tendem também a se afastar das grandes concentrações de oportunidades. Junta-se a isso o fato de, na maior parte dos casos, a oferta de serviços de transporte público de média e alta capacidade ser pior em locais com maior concentração de pessoas de baixa renda. Como consequência, seus níveis de acessibilidade são, em média, muito menores do que os dos mais ricos, como o gráfico deixa claro.

Part III

SEÇÃO 3: Dados de transporte público

Objetivo: O objetivo desta seção é apresentar (1) o que é a especificação GTFS de dados de transporte público; (2) onde baixar dados GTFS de cidades brasileiras e do mundo; e (3) como analisar e manipular dados de GTFS usando R.

Dados de transporte público são peças fundamentais no planejamento de transportes em geral, e em análises de acessibilidade em particular. Para serem usados de forma que se tenha segurança no resultado das análises, esses dados precisam ser confiáveis e de simples inspeção e interpretação.

Tentando satisfazer esses critérios, cada vez mais agências de transporte público, tomadores de decisão e pesquisadores têm buscado utilizar dados estruturados conforme especificações abertas e colaborativas - ou seja, cujo formato seja decidido por uma comunidade de atores interessados, incluindo partes que produzem esses dados (agências de transporte público, por exemplo) e que os consomem (pesquisadores, desenvolvedores de ferramentas de planejamento, etc.). Embora uma especificação aberta não necessariamente resolva o problema da qualidade e da confiabilidade dos dados por ela descritos, ela traz várias vantagens. O uso de um formato padrão de dados para transporte público permite o desenvolvimento e compartilhamento de ferramentas e programas computacionais para análise desses dados. Isso cria um campo comum de compartilhamento e aprendizado entre agências de transporte de diferentes cidades e países. Assim, um programa desenvolvido por agências de transporte para qualquer cidade no Brasil, Estados Unidos ou Japão pode ser facilmente utilizado para quaisquer cidades do mundo que também organizam seus dados naquele formato padrão. Além disso, quanto mais amplamente utilizado é esse formato padrão de dados, maior tende a ser tanto a confiabilidade da especificação em si quanto a facilidade de inspeção e interpretação dos seus dados, visto que múltiplos atores detêm o conhecimento necessário para tal.

A especificação de dados aberta e colaborativa mais amplamente utilizada no contexto do planejamento de transporte público é o formato GTFS, sigla para *General Transit Feed Specification* (Especificação Geral de Redes de Transporte Público, em tradução livre). Seus usos abrangem tanto o planejamento quanto a operação de sistemas de transporte público. Como visto no capítulo [Chapter 3](#), os dados de GTFS também são uma peça fundamental para calcular estimativas de acessibilidade urbana por transporte público. Nesta seção nós iremos aprender em mais detalhe o que são os dados GTFS, como eles são estruturados e como trabalhar com esses dados no R.

4 Dados GTFS

O formato GTFS, é uma especificação aberta e colaborativa que visa descrever os principais componentes de uma rede de transporte público. Atualmente, dados GTFS podem ser divididos em duas grandes categorias:

- *GTFS Schedule*, ou *GTFS Static*, que contém o cronograma estático de linhas de transporte público e informações espaciais sobre o itinerário de cada linha e suas paradas;
- *GTFS Realtime*, que contém informações de localização de veículos em tempo real e alertas de possíveis atrasos, de mudanças de percurso e de eventos que possam interferir no cronograma planejado.

Ao longo desta seção, nós focaremos no **GTFS Schedule**, por ser o formato de GTFS mais amplamente utilizado por agências de transporte e em análises de acessibilidade. Clique [aqui](#) para mais informações sobre o GTFS Realtime.

Por ser uma especificação aberta e colaborativa, o formato GTFS tenta abarcar em sua definição um grande número de usos distintos que agências de transporte e desenvolvedores de ferramentas possam dar a ele. No entanto, agências e softwares podem ainda assim depender de informações que não constem na especificação oficial. Surgem, dessa forma, extensões da especificação. Algumas dessas [extensões](#) podem eventualmente se tornar parte da especificação oficial, caso isto seja aceito pela comunidade de usuários do GTFS. Nesta seção nós focaremos em um subconjunto de informações presentes no formato GTFS Schedule “puro”, e, portanto, não cobriremos suas extensões.

4.1 Estrutura de arquivos GTFS

Usualmente, refere-se a um arquivo no formato GTFS Schedule (daqui em diante chamado apenas de GTFS) como *feed*. Neste livro, nós utilizaremos os termos *feed* e arquivo GTFS como sinônimos.

Um *feed* é nada mais do que um arquivo comprimido em formato `.zip` que contém um conjunto de tabelas salvas em formato `.txt` com algumas informações sobre a rede de transporte público (e.g. localização das paradas, frequências das viagens, traçado das rotas etc). Como em uma base de dados relacional, as tabelas de um *feed* têm algumas colunas-chave que permitem vincular os dados rotas, viagens, tabelas de horários etc. O esquema geral do GTFS é

mostrado na Figure 4.1, mostrando apenas algumas das principais tabelas e com as colunas-chave destacadas como os pontos finais das setas.

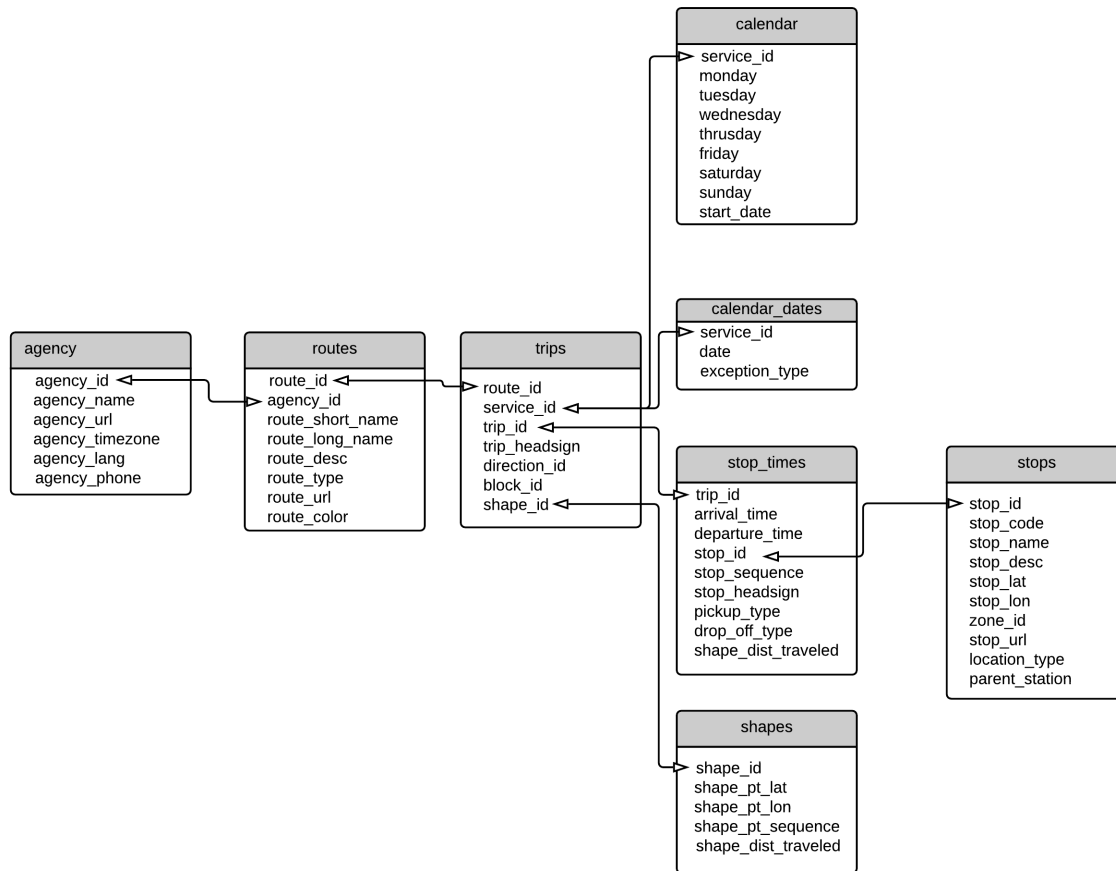


Figure 4.1: Esquema do formato GTFS

Fonte: (pereira2022gtfs2gps?)

Ao todo, 22 tabelas compõem o formato GTFS¹. Nem todas, no entanto, são obrigatórias para que um feed seja considerado válido - sendo consideradas, portanto, opcionais. A especificação classifica cada tabela conforme sua obrigatoriedade em três possíveis categorias: obrigatórias, opcionais e condicionalmente obrigatórias (quando a obrigatoriedade de uma tabela depende da existência de uma determinada tabela, coluna ou valor). Para fins de simplicidade, neste livro nós consideraremos apenas as duas primeiras categorias e faremos comentários quanto à obrigatoriedade de cada tabela quando apropriado. Desta forma, ficam assim classificadas as tabelas:

¹Disponível em <https://www.pasfor.com.br/>.

- **Obrigatórias:** `agency.txt`, `stops.txt`, `routes.txt`, `trips.txt`, `stop_times.txt`, `calendar.txt`.
- **Opcionais:** `calendar_dates.txt`, `fare_attributes.txt`, `fare_rules.txt`, `fare_products.txt`, `fare_leg_rules.txt`, `fare_transfer_rules.txt`, `areas.txt`, `stop_areas.txt`, `shapes.txt`, `frequencies.txt`, `transfers.txt`, `pathways.txt`, `levels.txt`, `translations.txt`, `feed_info.txt`, `attributions.txt`.

Ao longo desta seção, nós aprenderemos a estrutura básica de um arquivo GTFS e das tabelas que o compõem. Portanto, vamos olhar apenas para as tabelas obrigatórias e para as tabelas opcionais mais frequentemente utilizadas por produtores e consumidores desses arquivos. Para mais informações sobre as tabelas e as colunas não abordadas nesta seção, por favor verifique a [especificação oficial](#).

Nesta demonstração, nós utilizaremos um subconjunto de dados provenientes do *feed* da cidade de São Paulo criado pela SPTrans² e baixado em outubro de 2019. O *feed* contém as seis tabelas obrigatórias e mais duas tabelas opcionais bastante utilizadas, `shapes.txt` e `frequencies.txt`, o que permite uma boa visão geral sobre o formato GTFS.

4.1.1 `agency.txt`

Arquivo utilizado para descrever as operadoras de transporte que atuam no sistema descrito no arquivo GTFS. Embora o termo *agency* (agência) seja usado em lugar de *operators* (operadoras), por exemplo, fica a cargo do produtor do *feed* definir quais instituições serão listadas na tabela.

Por exemplo: múltiplas concessionárias de ônibus atuam em um determinado local, mas todo o planejamento de cronograma e de tarifa é realizado por uma instituição, em geral uma secretaria de transporte ou empresa pública específica. Esta instituição é também entendida pelos usuários do sistema como a operadora, de fato. Neste caso, devemos listar a instituição responsável pelo planejamento na tabela.

Agora, imagine um sistema em que a agência de transporte público local transfere a responsabilidade da operação de um sistema multimodal a diversas empresas, por meio de concessões. Cada uma dessas empresas é responsável pelo planejamento de cronogramas e tarifas dos modos que operam, desde que sejam seguidos determinados parâmetros pré-estabelecidos em contrato. Neste caso, deve-se listar as operadoras (concessionárias) na tabela, e não a agência de transporte público em si.

A [Table 4.1](#) mostra o arquivo `agency.txt` do *feed* da SPTrans. Como vocês podem ver, os responsáveis pelo *feed* optaram por listar a própria empresa no arquivo, e não as concessionárias que operam os ônibus e o metrô da cidade.

²Mais detalhes sobre o uso e a sintaxe do `data.table` podem ser lidos em <https://rdatatable.gitlab.io/data.table/index.html>.

Table 4.1: Exemplo de arquivo `agency.txt`

agency_id	agency_name	agency_url	agency_timezone	agency_lang
1	SPTRANS	http://www.sptrans.com.br/?versao=011019	America/Sao_Paulo	pt

É necessário notar que, embora estejamos apresentando o `agency.txt` em formato de tabela, o arquivo deve ser formatado como se fosse salvo em formato `.csv`. Ou seja, os valores de cada célula da tabela devem ser separados por vírgulas, e cada linha da tabela deve constar em uma linha no arquivo. A tabela acima, por exemplo, é definida da seguinte forma:

```
agency_id,agency_name,agency_url,agency_timezone,agency_lang
1,SPTRANS,http://www.sptrans.com.br/?versao=011019,America/Sao_Paulo,pt
```

Por uma questão de comunicação e interpretação dos dados, ao longo desta seção sempre apresentaremos os exemplos em formato de tabela. É importante ter em mente, porém, que essas tabelas são internamente organizadas no GTFS como mostrado acima.

4.1.2 stops.txt

Arquivo usado para descrever as paradas de transporte público que compõem o sistema. Os pontos listados neste arquivo podem fazer menção a paradas mais simples (como pontos de ônibus), estações, plataformas, entradas e saídas de estações, etc. A Table 4.2 mostra o `stops.txt` do *feed* da SPTrans.

Table 4.2: Exemplo de arquivo `stops.txt`

stop_id	stop_name	stop_desc	stop_lat	stop_lon
706325	Parada 14 Bis B/C	Viad. Dr. Plínio De Queiroz, 901	-	-
810602	R. Sta. Rita, 56	Ref.: R. Bresser / R. João Boemer	23.55593	46.65011
910776	Av. Do Estado, 5854	Ref.: Rua Dona Ana Néri	-	-
1010092	Parada Caetano Pinto	Av. Rangel Pestana, 1249 Ref.: Rua Caetano Pinto/rua Prof. Batista De Andrade	23.55896	46.61520
1010093	Parada Piratininga	Av. Rangel Pestana, 1479 Ref.: Rua Monsenhor Andrade	23.54615	46.62218
			23.54509	46.62006

stop_id	stop_name	stop_desc	stop_lat	stop_lon
1010099R.612	Xavantes,	Ref.: Rua Joli	-	-
			23.53545	46.61368

As colunas **stop_id** e **stop_name** servem como identificadores de cada parada, porém cumprem papéis distintos. O principal propósito da **stop_id** é identificar relações entre esta tabela e outras que compõem a especificação (como veremos mais à frente no arquivo **stop_times.txt**). Já a coluna **stop_name** serve como um identificador que seja facilmente reconhecido pelo usuário do sistema. Seus valores, portanto, costumam ser nomes de estações, nomes de pontos de interesse da cidade ou endereços (como no caso do *feed* da SPTrans).

A coluna **stop_desc**, presente no *feed* da SPTrans, é opcional e permite à agência de transporte adicionar alguma descrição de cada parada e do seu entorno para facilitar a sua identificação. As colunas **stop_lat** e **stop_lon**, por fim, são as responsáveis por associar cada parada a uma posição espacial, através de suas coordenadas geográficas de latitude e longitude.

Entre as colunas opcionais não presentes neste feed estão a **location_type** e a **parent_station**. A **location_type** é utilizada para denotar o tipo de localização a que cada ponto se refere. Quando ausente, todos os pontos são interpretados como paradas de transporte público, mas valores distintos podem ser usados para distinguir uma parada (**location_type** = 0) de uma estação (**location_type** = 1) ou uma área de embarque (**location_type** = 2), por exemplo. A coluna **parent_station** é utilizada para descrever relações de hierarquia entre dois pontos. Por exemplo, uma área de desembarque deve dizer a qual parada/plataforma ela pertence e uma parada/plataforma pode também, opcionalmente, listar a qual estação ela pertence.

4.1.3 routes.txt

Arquivo usado para descrever as linhas de transporte público que rodam no sistema descrito pelo arquivo GTFS, incluindo os modos de transporte utilizados em cada uma. A Table 4.3 mostra o **routes.txt** do *feed* da SPTrans.

Table 4.3: Exemplo de arquivo **routes.txt**

route_id	agency_id	route_short_name	route_long_name	route_type
CPTM L07	1	CPTM L07	JUNDIAI - LUZ	2
CPTM L08	1	CPTM L08	AMADOR BUENO - JULIO PRESTES	2
CPTM L09	1	CPTM L09	GRAJAU - OSASCO	2

route_id	agency_id	route_short_name	route_long_name	route_type
CPTM L10	1	CPTM L10	RIO GRANDE DA SERRA - BRÁS	2
CPTM L11	1	CPTM L11	ESTUDANTES - LUZ	2
CPTM L12	1	CPTM L12	CALMON VIANA - BRAS	2

Assim como no caso do arquivo `stops.txt`, a tabela do `routes.txt` também possui diferentes colunas que apontam o identificador de cada linha (`route_id`) e o seu nome. Neste caso, no entanto, existem duas colunas de nome: a `route_short_name` e a `route_long_name`. A primeira diz respeito ao nome da linha, usualmente utilizado por passageiros no dia-a-dia, enquanto o segundo tende a ser um nome mais descritivo. A SPTrans, por exemplo, optou por destacar os pontos finais de cada linha nesta coluna. Podemos notar também que os mesmos valores se repetem nas colunas `route_id` e `route_short_name`, o que não é obrigatório e nem proibido - neste caso, o produtor do *feed* julgou que os nomes das linhas poderiam funcionar satisfatoriamente como identificadores por serem razoavelmente curtos e não se repetirem.

A coluna `agency_id` é a chave que permite relacionar a tabela das rotas com a tabela descrita no `agency.txt`. Ela faz menção a uma agência descrita naquele arquivo, no caso a agência de id 1 (a própria SPTrans). Esta coluna é opcional no caso de *feeds* em que existe apenas uma agência, porém é obrigatória no caso em que existem mais de uma. Imaginemos, por exemplo, um *feed* que descreve um sistema multimodal que conta com um corredor de metrô e diversas linhas de ônibus: uma configuração possível de `routes.txt` descreveria as linhas de metrô como de responsabilidade da operadora do metrô, e as de ônibus como de responsabilidade da empresa responsável pelo planejamento das linhas de ônibus, por exemplo.

A coluna `route_type` é utilizada para descrever o modo de transporte utilizado em cada linha. Esta coluna aceita diferentes números, cada um representando um determinado modo. Este exemplo descreve linhas de trem, cujo valor numérico correspondente é 2. Os valores correspondentes para outros modos de transporte são listados na [especificação](#).

4.1.4 trips.txt

Arquivo usado para descrever as viagens realizadas pelo sistema de transporte público descrito pelo *feed*. A viagem é a unidade básica de movimento do formato GTFS: ela associa diferentes partidas de cada linha de transporte público a uma linha de transporte público (`route_id`), um serviço que opera em determinados dias da semana (como veremos mais à frente no arquivo `calendar.txt`) e uma trajetória espacial de viagem (como veremos mais à frente no arquivo `shapes.txt`). A Table 4.4 mostra o `trips.txt` do *feed* da SPTrans.

Table 4.4: Exemplo de arquivo `trips.txt`

trip_id	route_id	service_id	trip_headsign	direction_id	shape_id
CPTM L07-0	CPTM L07	USD	JUNDIAI	0	17846
CPTM L07-1	CPTM L07	USD	LUZ	1	17847
CPTM L08-0	CPTM L08	USD	AMADOR BUENO	0	17848
CPTM L08-1	CPTM L08	USD	JULIO PRESTES	1	17849
CPTM L09-0	CPTM L09	USD	GRAJAU	0	17850
CPTM L09-1	CPTM L09	USD	OSASCO	1	17851

A coluna `trip_id` identifica cada uma das viagens descritas na tabela. A `route_id` faz referência a uma linha de transporte público identificada no arquivo `routes.txt`. A coluna `service_id` identifica serviços que determinam os dias da semana em que cada uma das viagens opera (dias úteis, finais de semana, uma mistura dos dois, etc.), descritos detalhadamente no arquivo `calendar.txt`. A última coluna à direita na tabela acima é a `shape_id`, que identifica a trajetória espacial de cada uma das viagens, descrita em detalhes no arquivo `shapes.txt`.

As duas colunas restantes, `trip_headsign` e `direction_id`, são opcionais e devem ser utilizadas para descrever o sentido/destino da viagem. A primeira, `trip_headsign`, é utilizada para ditar o texto que aparece no letreiro de veículos (no caso de um ônibus, por exemplo) ou em painéis informativos (como em metrô e trens) destacando o destino da viagem. Já a coluna `direction_id` é frequentemente utilizada em conjunto com a primeira para dar uma conotação de ida ou volta para cada viagem, onde 0 representa ida e 1 volta, ou vice-versa (assim como ida e volta são conceitos que mudam conforme o referencial, os valores 0 e 1 podem ser usados como desejado, desde que um represente um sentido e o outro o contrário). No caso do nosso exemplo, as duas primeiras linhas são viagens que fazem menção à mesma rota de transporte público (CPTM L07), porém em sentidos opostos: uma corre em direção a Jundiaí, e a outra à Luz.

4.1.5 calendar.txt

Arquivo usado para descrever os diferentes tipos de serviço existentes no sistema de transporte público descrito pelo *feed*. Um serviço, neste contexto, denota um conjunto de dias da semana em que viagens são realizadas. Cada serviço também é definido pela data em que começa a

valer e pela data a partir do qual ele não é mais válido. A Table 4.5 mostra o `calendar.txt` do *feed* da SPTrans.

Table 4.5: Exemplo de arquivo `calendar.txt`

service_id	monday	tuesday	wednesday	thursday	friday	saturday	sunday	start_date	end_date
USD	1	1	1	1	1	1	1	20080101	20200501
U__	1	1	1	1	1	0	0	20080101	20200501
US_	1	1	1	1	1	1	0	20080101	20200501
_SD	0	0	0	0	0	1	1	20080101	20200501
__D	0	0	0	0	0	0	1	20080101	20200501
S	0	0	0	0	0	1	0	20080101	20200501

A coluna `service_id` identifica cada um dos serviços descritos na tabela. Como mostrado anteriormente, este identificador é usado também no arquivo `trips.txt`, e é o responsável por associar cada viagem a um determinado serviço.

As colunas `monday`, `tuesday`, `wednesday`, `thursday`, `friday`, `saturday` e `sunday` (segunda-feira a domingo, em inglês) são utilizadas para delimitar os dias em que cada serviço funciona. Um valor de 1 significa que o serviço opera em um determinado dia, enquanto um valor de 0 significa que ele não opera. Como podemos ver no exemplo acima, o serviço USD opera em todos os dias da semana. Já o serviço U__ opera apenas em dias úteis.

Por fim, as colunas `start_date` e `end_date` delimitam o intervalo em que cada serviço é válido. As datas do formato GTFS são sempre formatadas segundo a regra YYYYMMDD - ou seja, os primeiros quatro números definem o ano, os dois subsequentes definem o mês e os últimos dois, o dia. O valor 20220428, por exemplo, representa o dia 28 de abril de 2022.

4.1.6 shapes.txt

Arquivo usado para descrever a trajetória espacial de cada viagem listada no *feed*. Este arquivo é opcional, mas fortemente recomendado que agências de transporte o incluam em seus arquivos de GTFS. A Table 4.6 mostra o `shapes.txt` do *feed* da SPTrans.

Table 4.6: Exemplo de arquivo `shapes.txt`

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence
17846	-23.53517	-46.63535	1
17846	-23.53513	-46.63548	2
17846	-23.53494	-46.63626	3
17846	-23.53473	-46.63710	4
17846	-23.53466	-46.63735	5

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence
17846	-23.53416	-46.63866	6

A coluna **shape_id** identifica cada uma das trajetórias descritas na tabela. Como mostrado anteriormente, este identificador é usado também no arquivo **trips.txt**, e é o responsável por associar cada viagem à sua trajetória espacial. Diferentemente de todos os outros identificadores que vimos até então, no entanto, o identificador **shape_id** se repete em diversas observações da tabela. Isso porque o arquivo apresenta para cada **shape_id** uma série de pontos espaciais, cujas coordenadas geográficas são apresentadas nas colunas **shape_pt_lat** e **shape_pt_lon**. Por sua vez, a coluna **shape_pt_sequence** lista a sequência na qual cada ponto se conecta para formar a trajetória de cada **shape_id**. Os valores listados nesta coluna devem ser ordenados de forma crescente.

4.1.7 stop_times.txt

Arquivo usado para descrever a tabela de horários de cada viagem, incluindo o horário de chegada e partida em cada uma das paradas. A formatação deste arquivo depende da existência ou não de um arquivo **frequencies.txt**, o qual cobriremos a seguir. Por enquanto olharemos para o **stop_times.txt** do *feed* da SPTrans, que também conta com um **frequencies.txt**, na Table 4.7.

Table 4.7: Exemplo de arquivo **stop_times.txt**

trip_id	arrival_time	departure_time	stop_id	stop_sequence
CPTM L07-0	04:00:00	04:00:00	18940	1
CPTM L07-0	04:08:00	04:08:00	18920	2
CPTM L07-0	04:16:00	04:16:00	18919	3
CPTM L07-0	04:24:00	04:24:00	18917	4
CPTM L07-0	04:32:00	04:32:00	18916	5
CPTM L07-0	04:40:00	04:40:00	18965	6

A viagem cuja tabela de horários está sendo descrita é identificada pela coluna **trip_id**. De forma análoga ao que acontece na tabela de trajetórias, um mesmo **trip_id** se repete em muitas observações da tabela. Isso porque, assim como uma trajetória é composta por uma sequência de pontos espaciais, uma tabela de horários é composta por uma sequência de diversos horários de partida/chegada numa sequência de paradas de transporte público.

As colunas seguintes, **arrival_time**, **departure_time** e **stop_id**, são as responsáveis por descrever o cronograma de cada viagem, associando um horário de chegada e um horário de partida para cada uma das paradas visitadas na viagem. As colunas de horário são formatadas

segundo a regra HH:MM:SS - ou seja, os dois primeiros números definem a hora, os dois subsequentes os minutos e os últimos dois, os segundos. É importante ainda comentar que esta formatação aceita valores de hora maiores do que 24. Isto quer dizer que uma viagem cuja última parada seja visitada às 1h da manhã do dia seguinte ao que começou a rodar (digamos que ela tenha partido do ponto inicial às 23h) deve ter como horário de chegada 25:00:00, e não 01:00:00. A coluna `stop_id`, por sua vez, associa a parada visitada à uma descrição feita no arquivo `stops.txt`. Por fim, a coluna `stop_sequence` lista a sequência na qual cada parada se conecta às demais para formar o cronograma da viagem. Seus valores devem ser sempre ordenados de forma crescente.

Vale destacar aqui a diferença entre os arquivos `shapes.txt` e `stop_times.txt`. Embora os dois descrevam uma viagem espacialmente, eles o fazem de forma diferente. O `stop_times.txt` descreve a sequência de paradas e horários que compõem um cronograma, mas nada diz sobre o trajeto percorrido pelo veículo entre cada uma das paradas. Já o `shapes.txt` traz a trajetória detalhada da viagem como um todo, mas não descreve em que ponto do espaço estão as paradas da viagem. Quando usamos os dois arquivos em conjunto, portanto, sabemos tanto o cronograma de cada viagem quanto a trajetória espacial da viagem entre paradas.

4.1.8 frequencies.txt

Arquivo opcional usado para descrever a frequência de cada viagem dentro de um determinado período do dia. A Table 4.8 mostra o `frequencies.txt` do *feed* da SPTrans.

Table 4.8: Exemplo de arquivo `frequencies.txt`

trip_id	start_time	end_time	headway_secs
CPTM L07-0	04:00:00	04:59:00	720
CPTM L07-0	05:00:00	05:59:00	360
CPTM L07-0	06:00:00	06:59:00	360
CPTM L07-0	07:00:00	07:59:00	360
CPTM L07-0	08:00:00	08:59:00	360
CPTM L07-0	09:00:00	09:59:00	480

A viagem cuja frequência está sendo descrita é identificada pela coluna `trip_id`. Novamente, um mesmo id pode aparecer em várias observações da tabela. Isso porque a especificação GTFS prevê que uma mesma viagem pode ter frequências diferentes ao longo do dia (como em horários de pico e fora-pico, por exemplo). Assim, cada linha da tabela descreve um período do dia que se inicia no horário descrito na coluna `start_time` e termina no horário assinalado na coluna `end_time`.

Dentro do período especificado por essas duas colunas a viagem possui um *headway* detalhado na coluna `headway_secs`. O *headway* é o tempo que separa a passagem de dois veículos que

operam a mesma linha de transporte público. No caso desta tabela, esse tempo deve ser especificado em segundos. Um valor de 720 entre 04:00h e 05:00h, portanto, significa que a viagem CPTM L07-0 ocorre de 12 em 12 minutos dentro deste período.

Usando as tabelas `frequencies.txt` e `stop_times.txt` conjuntamente

É importante entender, agora, como a presença da tabela `frequencies.txt` altera a especificação da tabela `stop_times.txt`. Como podemos ver no nosso exemplo, o horário de partida definido no `stop_times.txt` para a viagem CPTM L07-0 é 04:00h, e ela chega na segunda parada às 04:08h. O cronograma de chegada e saída de uma mesma parada de uma viagem, no entanto, não pode ser definido mais de uma vez na tabela. Como então definir o cronograma das viagens que partem às 04:12h, 04:24h, 04:36h, etc. (lembre-se que o *headway* desta viagem é de 12 minutos)?

No caso em que a frequência de uma viagem é especificada no `frequencies.txt`, o cronograma (a tabela de horários) de uma viagem definido no `stop_times.txt` deve ser entendido como uma referência que descreve o tempo entre paradas. Isto é, os horários ali definidos não devem ser interpretados à risca. Por exemplo, o cronograma listado estabelece que o tempo de viagem entre a primeira e a segunda parada é de 8 minutos, e o tempo entre a segunda e a terceira também. Ou seja, a viagem que parte da primeira parada às 04:00h chega na segunda às 04:08h, e na terceira às 04:16h. A próxima viagem, que parte da primeira parada às 04:12h, por sua vez, chega na segunda parada às 04:20h, e na terceira às 04:28h.

Assumindo partidas de 12 em 12 minutos a partir das 04:00h, por exemplo, poderíamos descrever as mesmas viagens no `stop_times.txt` sem fazer uso do arquivo `frequencies.txt`. Para isso poderíamos adicionar um sufixo que identificasse cada uma das viagens referentes à linha (`route_id`) com identificador CPTM L07 ao longo do dia. A viagem (`trip_id`) com identificador CPTM L07-0_1, por exemplo, seria a primeira viagem do dia e partiria da primeira parada às 04:00h e chegaria na segunda parada às 04:08h. A viagem CPTM L07-0_2, por sua vez, seria a segunda viagem e partiria da primeira parada às 04:12h e chegaria na segunda às 04:20h, e daí em diante. Cada uma dessas viagens deveria ser também adicionada ao arquivo `trips.txt` e a quaisquer outros que possuam a coluna `trip_id` como identificador.

Outra variável que afeta a forma na qual o `frequencies.txt` afeta as tabelas de horários na tabela `stop_times.txt` é a coluna opcional `exact_times`. Um valor de 0 nesta coluna (ou quando ela está ausente do *feed*, como no caso do arquivo GTFS da SPTrans) indica que a viagem não necessariamente segue um cronograma fixo ao longo do período. Ao invés disso, operadores tentam se ater a um determinado *headway* durante o período. Naquele exemplo de uma viagem cujo *headway* é de 12 minutos entre as 04:00h e 05:00h, isto significaria que não necessariamente a primeira partida sairá exatamente às 04:00h, a segunda às 04:12h, e por aí em diante. A primeira pode, por exemplo, sair às 04:02h. A segunda, às 04:14h ou 04:13h, etc. Caso desejemos definir um cronograma que é seguido à risca, obtendo o mesmo resultado que seria obtido se definíssemos diversas viagens semelhantes partindo em diferentes horários no `stop_times.txt` (como mostrado anteriormente), devemos utilizar o valor 1 na coluna `exact_times`.

4.2 Onde encontrar GTFS de cidades brasileiras

No Brasil, existem dados de GTFS para diversas cidades. Em muitos casos, no entanto, esses dados são de propriedade de empresas operadoras e concessionárias de transporte, e não do poder público. Infelizmente, esses arquivos raramente são disponibilizados abertamente e publicamente, contrariando boas práticas de gestão e compartilhamento de dados de interesse público. A Table 4.9 mostra as fontes dos dados GTFS de algumas das poucas cidades do Brasil que disponibilizavam seus feeds abertamente no momento da publicação deste livro.

Table 4.9: Fontes de dados GTFS publicamente disponíveis no Brasil

Cidade	Fonte	Informações
Belo Horizonte	BHTrans	Dado aberto: transporte convencional ; transporte suplementar .
Fortaleza	ETUFOR	Dado aberto .
Fortaleza	Metrofor	Dado aberto .
Porto Alegre	EPTC	Dado aberto .
Rio de Janeiro	SMTR	Dado aberto .
São Paulo	EMTU	Download neste link . Necessário cadastro.
São Paulo	SPTrans	Download neste link . Necessário cadastro.

Obs. Os dados de GTFS disponibilizados pela secretaria de mobilidade urbana do Rio de Janeiro não incluem os dados dos sistemas de trens e metrô.

5 Manipulação e visualização de dados GTFS

Usualmente, arquivos GTFS provenientes de fontes oficiais são utilizados para desenvolver análises e pesquisas que possuem diversos elementos comuns. Visando facilitar a leitura, o processamento e a análise desses dados, a equipe do Projeto Acesso a Oportunidades vêm desenvolvendo o pacote de R `{gtfstools}` [(Herszenhut2022gtfstools?)]¹, que oferece diversas funções que facilitam a manipulação e a exploração de feeds.

Neste capítulo, nós iremos passar por algumas das funcionalidades mais frequentemente utilizadas do pacote. Para isso, vamos utilizar uma amostra do *feed* da SPTrans apresentado no capítulo anterior, disponível dentro do `{gtfstools}`.

5.1 Leitura e manipulação básica de arquivos GTFS

A leitura de arquivos GTFS com o `{gtfstools}` é feita com a função `read_gtfs()`, que recebe uma *string* com o caminho do arquivo. Após sua leitura, o *feed* é representado como uma lista de `data.tables`, uma versão de alta performance da classe `data.frame`. Ao longo deste capítulo, nós vamos nos referir a esta lista de tabelas como um **objeto GTFS**. Por padrão, a função lê todas as tabelas `.txt` do *feed*:

```
# carrega biblioteca
library(gtfstools)

# aponta string com endereço do arquivo
path <- system.file("extdata/spo_gtfs.zip", package = "gtfstools")

# leitura do gtfs
gtfs <- read_gtfs(path)

# consulta o nome das tabelas dentro da lista
names(gtfs)
```

```
[1] "agency"      "calendar"    "frequencies" "routes"      "shapes"
[6] "stop_times"  "stops"       "trips"
```

¹Disponível em <https://www.pasfor.com.br/>.

Como podemos ver, cada `data.table` dentro do objeto GTFS é nomeado de acordo com a tabela que ele representa, porém sem a extensão `.txt`. Isso nos permite selecionar e manipular cada uma das tabelas separadamente. O código abaixo, por exemplo, mostra os 6 primeiros registros da tabela `trips`:

```
head(gtfs$trips)
```

	route_id	service_id	trip_id	trip_headsign	direction_id	shape_id
1:	CPTM L07	USD	CPTM L07-0	JUNDIAI	0	17846
2:	CPTM L07	USD	CPTM L07-1	LUZ	1	17847
3:	CPTM L08	USD	CPTM L08-0	AMADOR BUENO	0	17848
4:	CPTM L08	USD	CPTM L08-1	JULIO PRESTES	1	17849
5:	CPTM L09	USD	CPTM L09-0	GRAJAU	0	17850
6:	CPTM L09	USD	CPTM L09-1	OSASCO	1	17851

As tabelas dentro de um objeto GTFS podem ser facilmente manipuladas usando a sintaxe de tabelas `data.table`. O pacote `data.table` oferece diversas funcionalidades úteis, como a edição de colunas por referência, filtros de linhas muito rápidos e agregação de dados eficiente². Para adicionar 100 segundos a todos os `headways` listados na tabela `frequencies` e reverter essa mudança em seguida, por exemplo, nós podemos usar o código abaixo:

```
# original
original_headway <- gtfs$frequencies$headway_secs
head(gtfs$frequencies, 3)
```

	trip_id	start_time	end_time	headway_secs
1:	CPTM L07-0	04:00:00	04:59:00	720
2:	CPTM L07-0	05:00:00	05:59:00	360
3:	CPTM L07-0	06:00:00	06:59:00	360

```
# frequência modificada
gtfs$frequencies[, headway_secs := headway_secs + 100]
head(gtfs$frequencies, 3)
```

	trip_id	start_time	end_time	headway_secs
1:	CPTM L07-0	04:00:00	04:59:00	820
2:	CPTM L07-0	05:00:00	05:59:00	460
3:	CPTM L07-0	06:00:00	06:59:00	460

²Mais detalhes sobre o uso e a sintaxe do `data.table` podem ser lidos em <https://rdatatable.gitlab.io/data.table/index.html>.

```
# recupera frequencia original
gtfs$frequencies[, headway_secs := original_headway]
head(gtfs$frequencies, 3)
```

	trip_id	start_time	end_time	headway_secs
1:	CPTM L07-0	04:00:00	04:59:00	720
2:	CPTM L07-0	05:00:00	05:59:00	360
3:	CPTM L07-0	06:00:00	06:59:00	360

Ao final de edições de um objeto GTFS no R, frequentemente vamos querer usar o GTFS manipulado para fazer análises de diferentes tipos. Para isso, é comum que precisemos do arquivo GTFS em formato `.zip` novamente, e não no formato de lista de tabelas dentro do R. O pacote `{gtfstools}` disponibiliza a função `write_gtfs()` exatamente com a finalidade de transformar objetos GTFS que existem apenas dentro do R em arquivos GTFS para arquivos compactados salvos no disco do seu computador. Para esta função, é necessário apenas listar o objeto e o endereço no qual o arquivo deve ser salvo:

```
# aponta endereço onde arquivo deve ser salvo
dest_path <- tempfile("new_gtfs", fileext = ".zip")

# salve GTFS localmente
write_gtfs(gtfs, dest_path)

# listar arquivos dentro do feed salvo
zip::zip_list(dest_path)[, c("filename", "compressed_size", "timestamp")]
```

	filename	compressed_size	timestamp
1	agency.txt	112	2022-10-11 22:25:20
2	calendar.txt	129	2022-10-11 22:25:20
3	frequencies.txt	2381	2022-10-11 22:25:20
4	routes.txt	659	2022-10-11 22:25:20
5	shapes.txt	160470	2022-10-11 22:25:20
6	stop_times.txt	7907	2022-10-11 22:25:20
7	stops.txt	18797	2022-10-11 22:25:20
8	trips.txt	717	2022-10-11 22:25:20

5.2 Cálculo de velocidade das linhas

Arquivos GTFS são frequentemente utilizados em estimativas de roteamento de transporte público e para informar passageiros sobre a tabela de horários das diferentes rotas que operam

em uma região. Dessa forma, é extremamente importante que o cronograma das viagens e a velocidade operacional de cada linha estejam adequadamente descritos no *feed*.

O `{gtfstools}` disponibiliza a função `get_trip_speed()` para facilitar o cálculo da velocidade de cada viagem presente no *feed*. Por padrão a função retorna a velocidade (em km/h) de todas as viagens do GTFS, mas viagens individuais também podem ser especificadas:

```
# calcular velocidade de todas viagens
speeds <- get_trip_speed(gtfs)
```

```
head(speeds)
```

	trip_id	origin_file	speed
1:	2002-10-0	shapes	8.952511
2:	2105-10-0	shapes	10.253365
3:	2105-10-1	shapes	9.795292
4:	2161-10-0	shapes	11.182534
5:	2161-10-1	shapes	11.784458
6:	4491-10-0	shapes	13.203560

```
nrow(speeds)
```

```
[1] 36
```

```
# calcular velocidade de duas viagens específicas
speeds <- get_trip_speed(gtfs, trip_id = c("CPTM L07-0", "2002-10-0"))
```

```
head(speeds)
```

	trip_id	origin_file	speed
1:	2002-10-0	shapes	8.952511
2:	CPTM L07-0	shapes	26.787768

Calcular a velocidade de uma viagem requer que nós saibamos o seu comprimento e em quanto tempo ela foi realizada. Para isso, portanto, a `get_trip_speed()` utiliza duas outras funções do `{gtfstools}` por trás dos panos: a `get_trip_length()` e a `get_trip_duration()`. O funcionamento das duas é muito parecido com o mostrado anteriormente, retornando o comprimento/duração de todas as viagens por padrão, ou de apenas algumas selecionadas, caso desejado. Abaixo nós mostramos seus comportamentos padrões:

```
# distância percorrida de todas viagens
length <- get_trip_length(gtfs, file = "shapes")

head(length)
```

```
      trip_id  length origin_file
1: CPTM L07-0 60.71894      shapes
2: CPTM L07-1 60.71894      shapes
3: CPTM L08-0 41.79037      shapes
4: CPTM L08-1 41.79037      shapes
5: CPTM L09-0 31.88906      shapes
6: CPTM L09-1 31.88906      shapes
```

```
# tempo de duração de todas viagens
duration <- get_trip_duration(gtfs)

head(duration)
```

```
      trip_id duration
1: 2002-10-0      48
2: 2105-10-0     108
3: 2105-10-1     111
4: 2161-10-0      94
5: 2161-10-1      93
6: 4491-10-0      69
```

Assim como a `get_trip_speed()` retorna velocidades em km/h por padrão, a `get_trip_length()` retorna os comprimentos em km e a `get_trip_duration()` retorna a duração em minutos. Essas unidades podem ser ajustadas com o argumento `unit`, presente em todas as funções.

5.3 Combinando e filtrando *feeds*

Muitas vezes o processo de processamento e edição de arquivos GTFS é realizado, em grande medida, manualmente. Consequentemente, pequenas inconsistências podem passar batidas pelos responsáveis por esse processamento. Um problema comumente observado em *feeds* é a presença de registros duplicados em uma mesma tabela. O *feed* da SPTrans, por exemplo, possui registros duplicados tanto no *agency.txt* quanto no *calendar.txt*:

```
gtfs$agency
```

```
agency_id agency_name agency_url
1:      1      SPTRANS http://www.sptrans.com.br/?versao=011019
2:      1      SPTRANS http://www.sptrans.com.br/?versao=011019
agency_timezone agency_lang
1: America/Sao_Paulo      pt
2: America/Sao_Paulo      pt
```

```
gtfs$calendar
```

```
service_id monday tuesday wednesday thursday friday saturday sunday
1:      USD      1      1      1      1      1      1      1
2:      U__      1      1      1      1      1      0      0
3:      US_      1      1      1      1      1      1      0
4:      _SD      0      0      0      0      0      1      1
5:      __D      0      0      0      0      0      0      1
6:      _S_      0      0      0      0      0      1      0
7:      USD      1      1      1      1      1      1      1
8:      U__      1      1      1      1      1      0      0
9:      US_      1      1      1      1      1      1      0
10:     _SD      0      0      0      0      0      1      1
11:     __D      0      0      0      0      0      0      1
12:     _S_      0      0      0      0      0      1      0
start_date  end_date
1: 2008-01-01 2020-05-01
2: 2008-01-01 2020-05-01
3: 2008-01-01 2020-05-01
4: 2008-01-01 2020-05-01
5: 2008-01-01 2020-05-01
6: 2008-01-01 2020-05-01
7: 2008-01-01 2020-05-01
8: 2008-01-01 2020-05-01
9: 2008-01-01 2020-05-01
10: 2008-01-01 2020-05-01
11: 2008-01-01 2020-05-01
12: 2008-01-01 2020-05-01
```

O {gtfstools} disponibiliza a função `remove_duplicates()` para remover essas duplicatas. Esta função recebe como *input* um objeto GTFS e retorna o mesmo objeto, porém sem registros duplicados:


```
# removendo valores duplicados
no_dups_gtfs <- remove_duplicates(gtfs)
```

```
# checagem do resultado
no_dups_gtfs$agency
```

```

  agency_id agency_name          agency_url
1:         1    SPTRANS http://www.sptrans.com.br/?versao=011019
  agency_timezone agency_lang
1: America/Sao_Paulo          pt

```

```
no_dups_gtfs$calendar
```

```

  service_id monday tuesday wednesday thursday friday saturday sunday
1:      USD      1       1           1           1       1           1
2:      U__      1       1           1           1       1           0
3:      US_      1       1           1           1       1           0
4:      _SD      0       0           0           0       0           1
5:      __D      0       0           0           0       0           1
6:      _S_      0       0           0           0       0           1
  start_date  end_date
1: 2008-01-01 2020-05-01
2: 2008-01-01 2020-05-01
3: 2008-01-01 2020-05-01
4: 2008-01-01 2020-05-01
5: 2008-01-01 2020-05-01
6: 2008-01-01 2020-05-01

```

Frequentemente, também, lidamos com múltiplos *feeds* em uma mesma área de estudo. Por exemplo, quando os dados dos sistemas de ônibus e de trens de uma mesma cidade estão salvos em arquivos GTFS separados. Neste caso, muitas vezes gostaríamos de uni-los em um único arquivo, diminuindo assim o esforço de manipulação e processamento dos dados. Para isso, o `{gtfstools}` disponibiliza a função `merge_gtfs()`. O exemplo abaixo mostra o resultado da combinação de dois *feeds* distintos, o da SPTrans (sem duplicatas) e o da EMTU, de Porto Alegre:

```
# leitura do GTFS de Porto Alegre
poa_path <- system.file("extdata/poa_gtfs.zip", package = "gtfstools")
poa_gtfs <- read_gtfs(poa_path)
```

```
poa_gtfs$agency
```

```
agency_id agency_name agency_url
1: EPTC Empresa Publica de Transportes e Circulação http://www.eptc.com.br
agency_timezone agency_lang agency_phone
1: America/Sao_Paulo pt 156
agency_fare_url
1: http://www2.portoalegre.rs.gov.br/eptc/default.php?p_secao=155
```

```
no_dups_gtfs$agency
```

```
agency_id agency_name agency_url
1: 1 SPTRANS http://www.sptrans.com.br/?versao=011019
agency_timezone agency_lang
1: America/Sao_Paulo pt
```

```
# Combinando GTFS de Porto Alegre e São Paulo
combined_gtfs <- merge_gtfs(no_dups_gtfs, poa_gtfs)
```

```
# checagem dos resultados
combined_gtfs$agency
```

```
agency_id agency_name
1: 1 SPTRANS
2: EPTC Empresa Publica de Transportes e Circulação
agency_url agency_timezone agency_lang
1: http://www.sptrans.com.br/?versao=011019 America/Sao_Paulo pt
2: http://www.eptc.com.br America/Sao_Paulo pt
agency_phone agency_fare_url
1:
2: 156 http://www2.portoalegre.rs.gov.br/eptc/default.php?p_secao=155
```

Como podemos ver, os registros das tabelas de ambos os *feeds* foram combinados em uma única tabela. Este é o caso quando os dois (ou mais, caso desejado) objetos GTFS possuem registros de uma mesma tabela (a *agency*, no exemplo). Caso apenas um dos objetos possua uma das tabelas, o resultado da operação de combinação copia esta tabela para o resultado final. É o caso, por exemplo, da tabela *frequencies*, que existe apenas no *feed* da SPTrans, mas não no da EMTU:

```
names(poa_gtfs)
```

```
[1] "agency"      "calendar"    "routes"      "shapes"      "stop_times"
[6] "stops"       "trips"
```

```
names(no_dups_gtfs)
```

```
[1] "agency"      "calendar"    "frequencies" "routes"      "shapes"
[6] "stop_times"  "stops"       "trips"
```

```
names(combined_gtfs)
```

```
[1] "agency"      "calendar"    "frequencies" "routes"      "shapes"
[6] "stop_times"  "stops"       "trips"
```

```
identical(no_dups_gtfs$frequencies, combined_gtfs$frequencies)
```

```
[1] TRUE
```

Um outro tipo de operação muito utilizada no tratamento de arquivos GTFS é o de filtragem desses arquivos. Frequentemente, *feeds* são usados para descrever redes de transporte público de grande escala, o que pode transformar sua edição, análise e transferência em operações mais complexas. Por esse motivo, pesquisadores e planejadores muitas vezes precisam trabalhar com um subconjunto de dados descritos nos *feeds*. Caso desejemos estimar a performance da rede de transporte em uma determinada região no horário de pico da manhã, por exemplo, podemos filtrar o nosso arquivo GTFS de modo a manter apenas os registros referentes a viagens que ocorrem nesse intervalo do dia.

O pacote `{gtfstools}` também traz diversas funções para facilitar a filtragem de arquivos GTFS. São elas:

- `filter_by_agency_id()`
- `filter_by_route_id()`
- `filter_by_service_id()`
- `filter_by_shape_id()`
- `filter_by_stop_id()`
- `filter_by_trip_id()`

- `filter_by_route_type()`
- `filter_by_weekday()`
- `filter_by_time_of_day()`
- `filter_by_sf()`

As seis primeiras (`filter_by_agency_id()`, `filter_by_route_id()`, `filter_by_service_id()`, `filter_by_shape_id()`, `filter_by_stop_id()` e `filter_by_trip_id()`) funcionam de forma muito similar. O usuário deve especificar um vetor de identificadores, e a função mantém no objeto GTFS apenas os registros referentes a esses identificadores. O exemplo abaixo demonstra essa funcionalidade com a `filter_by_trip_id()`:

```
# checando tamanho do feed antes do filtro
utils::object.size(gtfs)
```

864568 bytes

```
head(gtfs$trips[, .(trip_id, trip_headsign, shape_id)])
```

	trip_id	trip_headsign	shape_id
1:	CPTM L07-0	JUNDIAI	17846
2:	CPTM L07-1	LUZ	17847
3:	CPTM L08-0	AMADOR BUENO	17848
4:	CPTM L08-1	JULIO PRESTES	17849
5:	CPTM L09-0	GRAJAU	17850
6:	CPTM L09-1	OSASCO	17851

```
# filtro mantendo apenas duas viagens
smaller_gtfs <- filter_by_trip_id(gtfs, trip_id = c("CPTM L07-0", "CPTM L07-1"))

# checando tamanho do feed após o filtro
utils::object.size(smaller_gtfs)
```

71592 bytes

```
head(smaller_gtfs$trips[, .(trip_id, trip_headsign, shape_id)])
```

	trip_id	trip_headsign	shape_id
1:	CPTM L07-0	JUNDIAI	17846
2:	CPTM L07-1	LUZ	17847

```
unique(smaller_gtfs$shapes$shape_id)
```

```
[1] "17846" "17847"
```

O código acima mostra que a função não filtra apenas a tabela `trips`, mas também as outras tabelas que fazem referência aos identificadores especificados. Por exemplo, a trajetória das viagens CPTM L07-0 and CPTM L07-1 é descrita pelos `shape_ids` 17846 and 17847, respectivamente. Esses são, portanto, os únicos identificadores da tabela `shapes` mantidos no GTFS filtrado.

A função também pode funcionar com o comportamento diametralmente oposto: em vez de definirmos os identificadores cujos registros devem ser *mantidos* no *feed*, podemos especificar os identificadores que devem ser *retirados* dele. Para isso, usamos o argumento `keep` com valor `FALSE`:

```
# filtro removendo duas viagens do feed
smaller_gtfs <- filter_by_trip_id(
  gtfs,
  c("CPTM L07-0", "CPTM L07-1"),
  keep = FALSE
)

head(smaller_gtfs$trips[, .(trip_id, trip_headsign, shape_id)])
```

	trip_id	trip_headsign	shape_id
1:	CPTM L08-0	AMADOR BUENO	17848
2:	CPTM L08-1	JULIO PRESTES	17849
3:	CPTM L09-0	GRAJAU	17850
4:	CPTM L09-1	OSASCO	17851
5:	CPTM L10-0	RIO GRANDE DA SERRA	17852
6:	CPTM L10-1	BRÁS	17853

```
head(unique(smaller_gtfs$shapes$shape_id))
```

```
[1] "17848" "17849" "17850" "17851" "17852" "17853"
```

Como podemos ver, as viagens especificadas, bem como suas trajetórias, não estão presentes no GTFS filtrado. A mesma lógica aqui demonstrada com a `filter_by_trip_id()` é válida para

as funções que filtram objetos GTFS pelos identificadores `agency_id`, `route_id`, `service_id`, `shape_id`, `stop_id` e `route_type`.

Filtro por dia da semana

Outra operação que recorrentemente aparece em análises que envolvem dados GTFS é a de manter serviços que funcionem apenas em determinados horários do dia ou dias da semana. Para isso, o pacote disponibiliza as funções `filter_by_weekday()` e `filter_by_time_of_day()`.

A `filter_by_weekday()` recebe os dias da semana (em inglês) cujos serviços que neles operam devem ser mantidos. Adicionalmente, a função também inclui o argumento `combine`, que define como filtros de dois ou mais dias funcionam. Quando este recebe o valor `"and"`, apenas serviços que operam em todos os dias especificados são mantidos. Quando recebe o valor `"or"`, serviços que operam em pelo menos um dos dias são mantidos:

```
# filtro mantendo apenas serviços que operam no sábado e no domingo
smaller_gtfs <- filter_by_weekday(
  no_dups_gtfs,
  weekday = c("saturday", "sunday"),
  combine = "and"
)

smaller_gtfs$calendar[, c("service_id", "sunday", "saturday")]
```

	service_id	sunday	saturday
1:	USD	1	1
2:	_SD	1	1

```
# filtro mantendo apenas serviços que operam ou sábado ou no domingo
smaller_gtfs <- filter_by_weekday(
  no_dups_gtfs,
  weekday = c("sunday", "saturday"),
  combine = "or"
)

smaller_gtfs$calendar[, c("service_id", "sunday", "saturday")]
```

	service_id	sunday	saturday
1:	USD	1	1
2:	US_	0	1
3:	_SD	1	1
4:	__D	1	0
5:	_S_	0	1

Filtro por hora do dia

A `filter_by_time_of_day()`, por sua vez, recebe o começo e o final de uma janela de tempo e mantém os registros relacionados a viagens que rodam dentro dessa janela de horário. O funcionamento da função depende da presença ou não da tabela `frequencies` no GTFS: o cronograma descrito na `stop_times` das viagens descritas na tabela `frequencies` não deve ser filtrado, pois, como comentado no capítulo anterior, ele serve como um modelo que dita o tempo de viagem entre uma parada e outra. Caso a `frequencies` esteja ausente, no entanto, a `stop_times` é filtrada segundo o intervalo de tempo especificado. Vamos ver como isso funciona com um exemplo:

```
# filtro mantendo apenas viagens dentro do período de 5 às 6 da manhã
smaller_gtfs <- filter_by_time_of_day(gtfs, from = "05:00:00", to = "06:00:00")

head(smaller_gtfs$frequencies)
```

	trip_id	start_time	end_time	headway_secs
1:	CPTM L07-0	05:00:00	05:59:00	360
2:	CPTM L07-1	05:00:00	05:59:00	360
3:	CPTM L08-0	05:00:00	05:59:00	480
4:	CPTM L08-1	05:00:00	05:59:00	480
5:	CPTM L09-0	05:00:00	05:59:00	480
6:	CPTM L09-1	05:00:00	05:59:00	480

```
head(smaller_gtfs$stop_times[, c("trip_id", "departure_time", "arrival_time")])
```

	trip_id	departure_time	arrival_time
1:	CPTM L07-0	04:00:00	04:00:00
2:	CPTM L07-0	04:08:00	04:08:00
3:	CPTM L07-0	04:16:00	04:16:00
4:	CPTM L07-0	04:24:00	04:24:00
5:	CPTM L07-0	04:32:00	04:32:00
6:	CPTM L07-0	04:40:00	04:40:00

```
frequencies <- gtfs$frequencies
gtfs$frequencies <- NULL

smaller_gtfs <- filter_by_time_of_day(gtfs, from = "05:00:00", to = "06:00:00")

head(smaller_gtfs$stop_times[, c("trip_id", "departure_time", "arrival_time")])
```

	trip_id	departure_time	arrival_time
1:	CPTM L07-0	05:04:00	05:04:00
2:	CPTM L07-0	05:12:00	05:12:00
3:	CPTM L07-0	05:20:00	05:20:00
4:	CPTM L07-0	05:28:00	05:28:00
5:	CPTM L07-0	05:36:00	05:36:00
6:	CPTM L07-0	05:44:00	05:44:00

O filtro da tabela `stop_times` pode funcionar de duas formas distintas. Uma opção é manter intactas todas as viagens que cruzam a janela de tempo especificada. A outra é manter apenas os segmentos de viagens que ocorrem dentro da janela (comportamento padrão da função). Este comportamento é controlado com o parâmetro `full_trips`, como mostrado a seguir (prestem atenção nos horários e nos segmentos presentes em cada exemplo):

```
# filtro mantendo apenas viagens inteiramente dentro do período de 5 às 6 da manhã
smaller_gtfs <- filter_by_time_of_day(
  gtfs,
  "05:00:00",
  "06:00:00",
  full_trips = TRUE
)

head(
  smaller_gtfs$stop_times[
    ,
    c("trip_id", "departure_time", "arrival_time", "stop_sequence")
  ]
)
```

	trip_id	departure_time	arrival_time	stop_sequence
1:	CPTM L07-0	04:00:00	04:00:00	1
2:	CPTM L07-0	04:08:00	04:08:00	2
3:	CPTM L07-0	04:16:00	04:16:00	3
4:	CPTM L07-0	04:24:00	04:24:00	4
5:	CPTM L07-0	04:32:00	04:32:00	5
6:	CPTM L07-0	04:40:00	04:40:00	6

```
# filtro mantendo apenas viagens que do período que estão funcionando entre 5 e 6 da manhã
smaller_gtfs <- filter_by_time_of_day(
  gtfs,
  "05:00:00",
```



```

    "06:00:00",
    full_trips = FALSE
  )

  head(
    smaller_gtfs$stop_times[
      ,
      c("trip_id", "departure_time", "arrival_time", "stop_sequence")
    ]
  )
)

```

	trip_id	departure_time	arrival_time	stop_sequence
1:	CPTM L07-0	05:04:00	05:04:00	9
2:	CPTM L07-0	05:12:00	05:12:00	10
3:	CPTM L07-0	05:20:00	05:20:00	11
4:	CPTM L07-0	05:28:00	05:28:00	12
5:	CPTM L07-0	05:36:00	05:36:00	13
6:	CPTM L07-0	05:44:00	05:44:00	14

Filtro espacial

Por fim, o pacote `{gtfstools}` também disponibiliza uma função que permite filtrar o objeto GTFS usando um polígono espacial. A `filter_by_sf()` recebe um objeto do tipo `sf/sfc` (representação espacial estabelecida pelo pacote `sf`), ou sua *bounding box*, e mantém os registros cujas viagens são selecionadas por uma operação espacial especificada pelo usuário. Embora aparentemente complicado, este processo de filtragem é muito facilmente compreendido quando apresentado visualmente. Para isso, vamos criar uma função auxiliar:

```

# carreg biblioteca ggplot2 para visualização de dados
library(ggplot2)

# cria uma função genérica para filtrar GTFS e plotar figura
plotter <- function(gtfs,
  geom,
  spatial_operation = sf::st_intersects,
  keep = TRUE,
  do_filter = TRUE) {
  if (do_filter) {
    gtfs <- filter_by_sf(gtfs, geom, spatial_operation, keep)
  }

  shapes <- convert_shapes_to_sf(gtfs)
}

```

```
trips <- get_trip_geometry(gtfs, file = "stop_times")
geom <- sf::st_as_sfc(geom)

ggplot() +
  geom_sf(data = trips) +
  geom_sf(data = shapes) +
  geom_sf(data = geom, fill = NA)
}
```

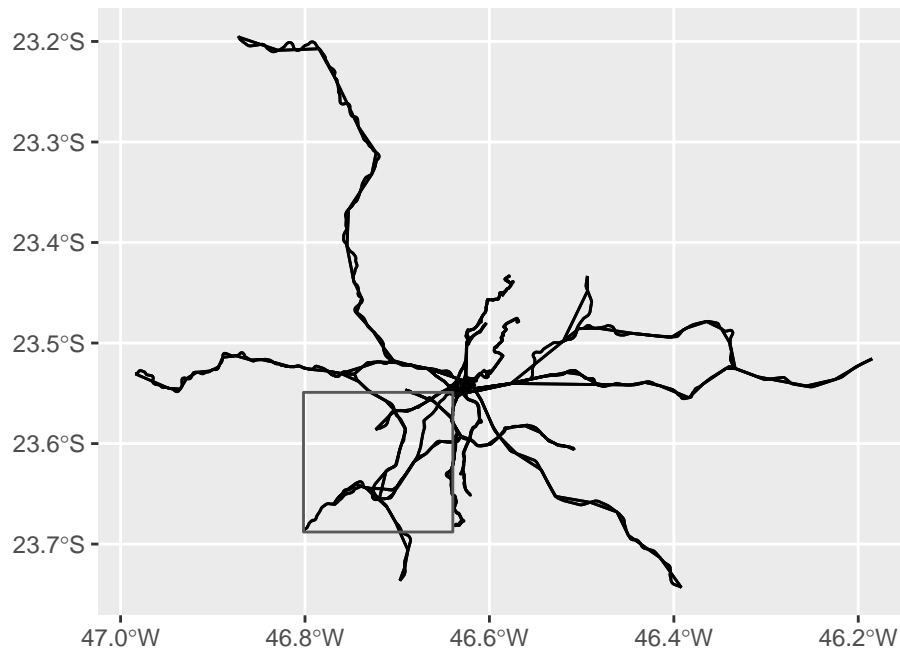
Esta função genérica criada acima:

- Possui os mesmos parâmetros da `filter_by_sf()`, com a exceção do parâmetro `do_filter`, usado para mostrar os dados não filtrados;
- Condicionalmente filtra um objeto GTFS usando uma *bounding box* chamada `geom`;
- Gera a geometria das viagens usando as funções `convert_shapes_to_sf()` e `get_trip_geometry()`;
- Cria um polígono a partir da *bounding box*;

Vamos usar como exemplo a filtragem usando a *bounding box* da trajetória de *id* 68962. O código abaixo apresenta a distribuição espacial dos dados não filtrados:

```
# converte para objeto espacial sf o traçado de um shape_id
# pega o bounding box desse traçado
bbox <- convert_shapes_to_sf(gtfs, shape_id = "68962") |>
  sf::st_bbox()

# plota toda rede de transporte público e o bounding box construído acima
plotter(gtfs, bbox, do_filter = FALSE)
```



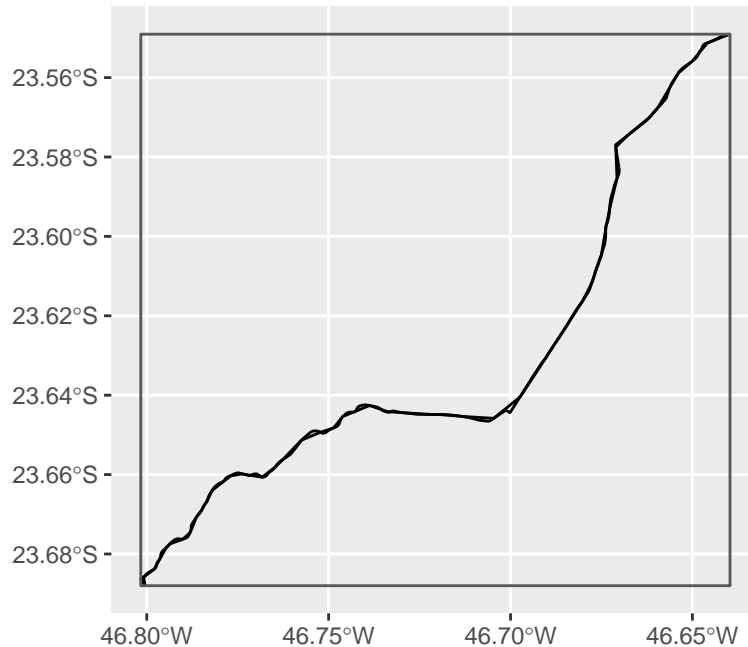
Note aqui que nós usamos a função `convert_shapes_to_sf()`, que converte uma determinada trajetória descrita no GTFS em um objeto espacial do tipo `sf`, para obter a *bounding box* da trajetória. Por padrão, a `filter_by_sf()` (e a `plotter()`, conseqüentemente) mantém os dados relacionados aos registros de viagens cujas trajetórias possuem alguma interseção com o polígono espacial selecionado:

```
plotter(gtfs, bbox)
```



Nós podemos, no entanto, controlar a operação espacial usada no processo de filtragem. Por exemplo, o código abaixo mostra como nós podemos manter os dados relacionados a viagens que estão *contidas* dentro do polígono espacial:

```
plotter(gtfs, bbox, spatial_operation = sf::st_contains)
```



5.4 Mapeamento do *headway* das linhas

Como mostrado nas seções anteriores, o `{gtfstools}` disponibiliza uma grande caixa de ferramentas que podem ser usadas no processamento e na análise de arquivos GTFS. O pacote, no entanto, oferece diversas outras funções que não puderam ser apresentadas neste livro, por questões de espaço. A lista completa de funções disponíveis no pacote pode ser conferida no [site do {gtfstools}](#).

A apresentação das funções feitas até aqui tem um importante caráter demonstrativo, porém não mostra como elas podem ser usadas de forma conjunta no desenvolvimento de uma análise de um arquivo GTFS. Esta seção preenche esta lacuna, mostrando como o pacote pode ser usado, por exemplo, para responder a seguinte pergunta: como se distribuem espacialmente os tempos entre veículos de uma mesma linha (ou seja, o *headway*) no GTFS da SPTrans?

A primeira etapa é definir o escopo da nossa análise. A fim de exemplo, vamos considerar o *headway* no pico da manhã, entre 7h e 9h, em uma típica terça-feira de operação. Para isso, precisamos filtrar o nosso *feed*:

```
# ler dados de GTFS
gtfs <- read_gtfs(path)

# filtrando GTFS
```

```

filtered_gtfs <- gtfs |>
  remove_duplicates() |>
  filter_by_weekday("tuesday") |>
  filter_by_time_of_day(from = "07:00:00", to = "09:00:00")

# checando resultado do filtro
filtered_gtfs$frequencies[trip_id == "2105-10-0"]

```

	trip_id	start_time	end_time	headway_secs
1:	2105-10-0	07:00:00	07:59:00	900
2:	2105-10-0	08:00:00	08:59:00	1200

```

# checando resultado do filtro
filtered_gtfs$calendar

```

	service_id	monday	tuesday	wednesday	thursday	friday	saturday	sunday
1:	USD	1	1	1	1	1	1	1
2:	U__	1	1	1	1	1	0	0

	start_date	end_date
1:	2008-01-01	2020-05-01
2:	2008-01-01	2020-05-01

Em seguida, precisamos calcular o *headway* dentro do período estabelecido. Essa informação pode ser encontrada na tabela *frequencies*, porém há um elemento complicador: cada viagem está associada a mais de um *headway*, como podemos ver acima (um entre 7:00 e 7:59, e outro entre 8:00 e 8:59). Para resolver esta questão, portanto, vamos calcular o *headway médio* neste intervalo entre 7:00 e 8:59.

Os primeiros registros da tabela *frequencies* do GTFS da SPTrans parecem sugerir que os períodos do dia estão listados sempre de uma em uma hora, porém isto não é uma regra estabelecida na especificação GTFS e nem é a prática adotada em outros GTFS. Por isso, nós vamos calcular a *média ponderada* do *headway* no período especificado. Para isso, nós precisamos multiplicar cada *headway* pelo intervalo de tempo em que ele é válido, e dividir o total desta soma pelo intervalo de tempo total (duas horas). Para calcular o intervalo de tempo em que cada *headway* é válido, nós usamos a função `convert_time_to_seconds()` para calcular o começo e o fim do intervalo em segundos e subtraímos o valor do fim pelo do começo, como abaixo:

```

filtered_gtfs <- convert_time_to_seconds(filtered_gtfs)

```

```
filtered_gtfs$frequencies[trip_id == "2105-10-0"]
```

	trip_id	start_time	end_time	headway_secs	start_time_secs	end_time_secs
1:	2105-10-0	07:00:00	07:59:00	900	25200	28740
2:	2105-10-0	08:00:00	08:59:00	1200	28800	32340

```
filtered_gtfs$frequencies[, time_interval := end_time_secs - start_time_secs]
```

Em seguida, nós calculamos o *headway* médio:

```
avg_headway <- filtered_gtfs$frequencies[
  ,
  .(avg_headway = weighted.mean(headway_secs, w = time_interval)),
  by = trip_id
]

avg_headway[trip_id == "2105-10-0"]
```

	trip_id	avg_headway
1:	2105-10-0	1050

```
head(avg_headway)
```

	trip_id	avg_headway
1:	CPTM L07-0	360
2:	CPTM L07-1	360
3:	CPTM L08-0	300
4:	CPTM L08-1	300
5:	CPTM L09-0	240
6:	CPTM L09-1	240

Nós precisamos agora gerar a trajetória espacial de cada viagem e juntar esta informação à do *headway* médio. Para isso, nós vamos utilizar a função `get_trip_geometry()`, que, dado um objeto GTFS, retorna a trajetória espacial de suas viagens. Esta função nos permite especificar de quais viagens nós queremos as trajetórias, logo nós vamos calcular apenas as daquelas que estão presentes na tabela de *headways* médios:

```

selected_trips <- avg_headway$trip_id

geoms <- get_trip_geometry(
  filtered_gtfs,
  trip_id = selected_trips,
  file = "shapes"
)

head(geoms)

```

Simple feature collection with 6 features and 2 fields

Geometry type: LINESTRING

Dimension: XY

Bounding box: xmin: -46.98404 ymin: -23.73644 xmax: -46.63535 ymax: -23.19474

Geodetic CRS: WGS 84

	trip_id	origin_file	geometry
1	CPTM L07-0	shapes	LINESTRING (-46.63535 -23.5...
2	CPTM L07-1	shapes	LINESTRING (-46.87255 -23.1...
3	CPTM L08-0	shapes	LINESTRING (-46.64073 -23.5...
4	CPTM L08-1	shapes	LINESTRING (-46.98404 -23.5...
5	CPTM L09-0	shapes	LINESTRING (-46.77604 -23.5...
6	CPTM L09-1	shapes	LINESTRING (-46.69711 -23.7...

O objeto `geoms` está no formato `sf`, e não no `data.table`, que precisamos que ele esteja para juntarmos à tabela de *headways*. Depois de convertê-lo para o formato desejado adequado e juntá-lo à tabela de *headways*, nós precisamos apenas configurar o nosso mapa como desejado. No exemplo abaixo, nós usamos cores e espessuras de linhas que variam de acordo com o *headway* de cada viagem:

```

library(data.table)
library(sf)

# converte objeto em data.table e faz 'merge' para fazer as infos espaciais
setDT(geoms)
avg_headway[geoms, on = "trip_id", geometry := i.geometry]

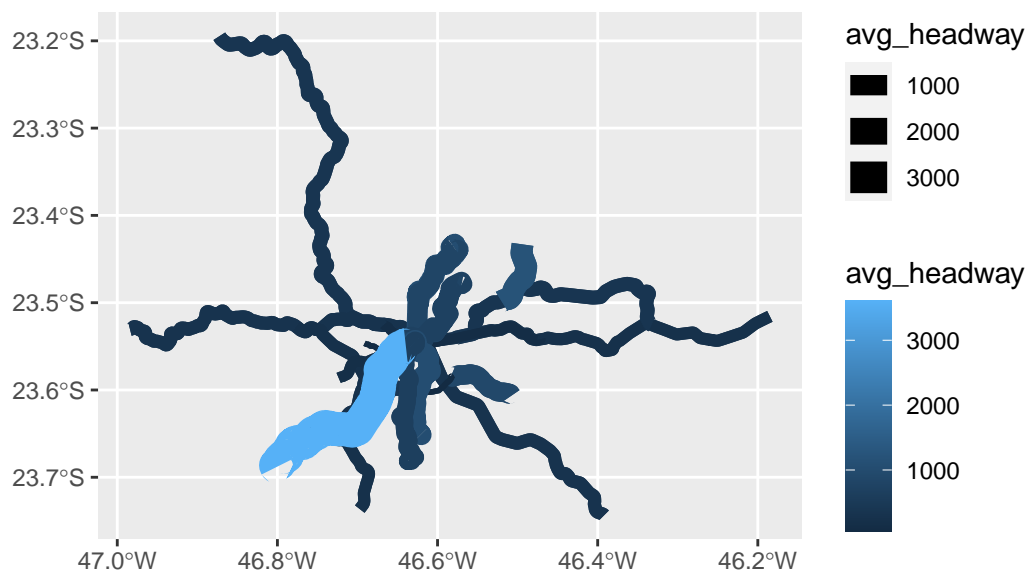
# converte objeto de volta para sf para poder fazer o mapa
avg_headway <- st_sf(avg_headway)

# plotar figura
ggplot(avg_headway) +

```



```
geom_sf(aes(color = avg_headway, size = avg_headway))
```



Como podemos ver, o pacote `{gtfstools}` torna o desenvolvimento de análises de *feeds* de transporte público algo fácil e que requer apenas o conhecimento básico de pacotes de manipulação de tabela (como o `data.table` ou `dplyr`) para grande parte das etapas que as compõem. O exemplo apresentado nesta seção mostra como muitas de suas funções podem ser usadas conjuntamente para revelar aspectos importantes de sistemas de transporte público descritos no formato GTFS.

Part IV

Objetivo: O objetivo desta seção é apresentar um exemplo de avaliação do impacto de um projeto de infra-estrutura de transportes sobre acessibilidade urbana usando R.

Embora muito utilizado na literatura científica, o foco em questões de acessibilidade urbana ainda é relativamente recente no planejamento e operação de sistemas de transporte público por parte de agências de transportes e tomadores de decisão (**boisjoly2017how?**). Muito disto se dá pela dificuldade de incorporar análises de acessibilidade a métodos de avaliação de projetos e à atividades de planejamento do dia-a-dia (**silva2017accessibility?**; **buttner2021accessibility?**).

Nesta seção, nós usamos um estudo de caso de um projeto de expansão de metrô para ilustrar podemos usar o R para avaliar o impacto de projetos de infraestrutura de transporte sobre acessibilidade urbana. O capítulo apresenta um método para medir tanto o impacto de investimentos de transporte sobre o nível médio de acessibilidade da população, como também para investigar o impacto de projetos sobre as desigualdades de acesso a oportunidades. A aplicação do método envolve a edição de arquivos GTFS, o cálculo de matrizes de custo, a tomada de decisão por trás da escolha de qual medida de acessibilidade utilizar, a estimativa dos níveis de acessibilidade, a visualização espacial desses níveis e o cálculo e análise de indicadores de desigualdade. Assim, esse estudo de caso engloba muitos dos pontos abordados neste livro, e serve como uma aplicação prática dos conceitos até então apresentados.

É importante mencionar que uma avaliação de projeto, investimento ou política pública deve abranger um grande leque de critérios de diferentes naturezas. Esses critérios vão desde aspectos de participação social na formulação da política/projeto até o seu impacto considerando dimensões ambientais, econômicas, sociais, entre outras. Embora uma avaliação de impacto de acessibilidade seja muito importante para a caracterização dos potenciais benefícios e desempenho de redes de transporte como um todo, ela é apenas uma dessas dimensões, e deve complementar ou ser complementada por outras análises.

6 Comparando a acessibilidade entre dois cenários de transporte

Neste capítulo, nós vamos fazer uma avaliação do projeto de expansão da Linha Leste do metrô de Fortaleza (LINK para FIGURA). O traçado da Linha Leste possui 7.3 km de extensão e liga o Centro de Fortaleza ao bairro Papicu, permitindo a integração das linhas de metrô Sul e Oeste com corredores de Veículo Leve sobre Trilhos (VLT) e o terminal de ônibus no Papicu. No momento em que esse livro foi escrito, o metrô ainda estava em fase de construção. Portanto, estamos tratando de uma avaliação *ex-ante* em que avaliamos o futuro impacto esperado desta expansão de infraestrutura sobre a acessibilidade urbana da cidade.

É importante notar que, na prática, a implementação deste investimento também deverá envolver mudanças nas frequências das linhas de metrô Sul e Oeste, além do racionamento do sistema de ônibus municipais, conforme o Plano de Acessibilidade Sustentável de Fortaleza (Pasfor)¹. Essas mudanças adicionais no sistema de transporte público não foram incorporados aos nossos cenários. Portanto, o estudo de caso que apresentamos aqui considera um cenário simplificado que considera apenas a implementação da Linha Leste. Para uma avaliação mais completa do impacto de acessibilidade da implementação da Linha Leste e das mudanças previstas no Pasfor, por favor veja (ref?).

Para a avaliação de impacto da Linha Leste do metrô, nós precisaremos comparar os níveis de acessibilidade antes e depois da implementação do novo corredor. Precisamos, portanto, utilizar diferentes arquivos GTFS para representar os cenários antes e depois, e seguir os seguintes passos: 1. Calcular duas matrizes de tempo de viagem, uma antes e outra depois do investimento 2. Calcular dois cenários de acessibilidade, um antes e outro depois do investimento 3. Comparar esses as condições de acessibilidade nos cenários antes e depois, examinando como os impactos de acessibilidade se distribuem espacialmente e entre populações de diferentes níveis socioeconômicos.

Neste capítulo, vamos cobrir este passo-a-passo em detalhes, começando primeiro pelo detalhamento dos GTFS que vamos usar.

¹Disponível em <https://www.pasfor.com.br/>.

6.1 Arquivos GTFS usados na análise

Nesta análise, usaremos os arquivos GTFS da METROFOR e da ETUFOR, que descrevem o sistema de transporte público operante na cidade de Fortaleza. Como esses são arquivos grandes, somando quase 20 MB no total, eles não estão disponíveis dentro dos pacotes `{r5r}` e `{gtfstools}`, como até então era o caso das amostras de *feeds* que usamos nos exemplos. Precisamos, portanto, baixá-los para o nosso projeto:

```
metrofor_path <- tempfile("metrofor", fileext = ".zip")
etufor_path <- tempfile("etufor", fileext = ".zip")

httr::GET(
  "https://github.com/ipeaGIT/intro_access_book/releases/download/data_1st_edition/gtfs_fo
  httr::write_disk(metrofor_path)
)

httr::GET(
  "https://github.com/ipeaGIT/intro_access_book/releases/download/data_1st_edition/gtfs_fo
  httr::write_disk(etufor_path)
)
```

Para simularmos a operação da nova Linha Leste do metrô, precisamos também de um *feed* que descreva o funcionamento deste corredor. Este GTFS foi criado para o estudo de (**ref?**), que avalia o impacto de acessibilidade da implementação da Linha Leste, e será reutilizado neste exemplo. Assim como os *feeds* da ETUFOR e da METROFOR, este arquivo GTFS está disponível para *download* no repositório do livro no GitHub:

```
linha_leste_path <- tempfile("linha_leste", fileext = ".zip")

httr::GET(
  "https://github.com/ipeaGIT/intro_access_book/releases/download/data_1st_edition/gtfs_li
  httr::write_disk(linha_leste_path)
)
```

Esses três arquivos GTFS serão usados em conjunto para calcular a acessibilidade dos cenários antes e depois. No primeiro, apenas os *feeds* da METROFOR e da ETUFOR serão usados no roteamento de transporte público, representando a típica operação de transporte público antes da implementação do novo corredor. No cenário pós-implementação, incluiremos também o GTFS da Linha Leste, incorporando à análise a operação planejada desta linha após sua finalização.

6.2 Cálculo das matrizes de tempo de viagem

Tendo definido quais arquivos GTFS vamos usar em cada um dos cenários analisados, precisamos agora utilizá-los para calcular as matrizes de tempo de viagem, que por sua vez vamos usar para estimar os níveis de acessibilidade. Como apresentado na [Section 3.1](#), para isso vamos utilizar a função `travel_time_matrix()`, do pacote `{r5r}`.

Antes de calcular as matrizes de fato, nós precisamos organizar os nossos arquivos na estrutura que o `{r5r}` requer. Cada um dos nossos cenários (antes e depois) deve ser representado por uma pasta contendo os arquivos necessários para o roteamento. Vamos criar uma pasta dentro da qual vamos botar as pastas do roteamento:

```
analysis_dir <- "impact_analysis"

dir.create(analysis_dir)
```

Dentro desta pasta, vamos criar as pastas de roteamento dos cenários antes e depois, contendo seus respectivos *feeds*:

```
dir_before <- file.path(analysis_dir, "before")
dir_after <- file.path(analysis_dir, "after")

dir.create(dir_before)
dir.create(dir_after)

file.copy(from = metrofor_path, to = file.path(dir_before, "metrofor.zip"))
file.copy(from = metrofor_path, to = file.path(dir_after, "metrofor.zip"))

file.copy(from = etufor_path, to = file.path(dir_before, "etufor.zip"))
file.copy(from = etufor_path, to = file.path(dir_after, "etufor.zip"))

file.copy(from = linha_leste_path, to = file.path(dir_after, "linha_leste.zip"))

fs::dir_tree(analysis_dir)
```

```
impact_analysis
+-- after
|   +-- etufor.zip
|   +-- linha_leste.zip
|   \-- metrofor.zip
\-- before
    +-- etufor.zip
```

```
\-- metrofor.zip
```

Para estimarmos o tempo de viagem na nossa área de estudo, precisamos ainda de um arquivo representando a rede viária do local, em formato `.pbf`, e opcionalmente iremos utilizar um arquivo representando a topografia local, em formato `.tif`. Esses arquivos, assim como os GTFS, estão disponíveis para *download* no repositório do livro. Partindo do pressuposto que a implementação da Linha Leste não afetaria o traçado das ruas e calçadas na região, bem como a topografia local, podemos usar os mesmos arquivos nas duas matrizes de tempo de viagem:

```
pbf_path <- tempfile("street_network", fileext = ".osm.pbf")
tif_path <- tempfile("topography", fileext = ".tif")

httr::GET(
  "https://github.com/ipeaGIT/intro_access_book/releases/download/data_1st_edition/for_202
  httr::write_disk(pbf_path)
)

httr::GET(
  "https://github.com/ipeaGIT/intro_access_book/releases/download/data_1st_edition/topogra
  httr::write_disk(tif_path)
)

file.copy(from = pbf_path, to = file.path(dir_before, "street_network.osm.pbf"))
file.copy(from = pbf_path, to = file.path(dir_after, "street_network.osm.pbf"))

file.copy(from = tif_path, to = file.path(dir_before, "topography.tif"))
file.copy(from = tif_path, to = file.path(dir_after, "topography.tif"))

fs::dir_tree(analysis_dir)
```

```
impact_analysis
+-- after
|   +-- etufor.zip
|   +-- linha_leste.zip
|   +-- metrofor.zip
|   +-- street_network.osm.pbf
|   \-- topography.tif
\-- before
    +-- etufor.zip
    +-- metrofor.zip
```

```
+-- street_network.osm.pbf
\-- topography.tif
```

Finalmente, podemos começar o cálculo das matrizes. A primeira etapa é construir a rede de transporte multimodal usada pelo {r5r} no roteamento e criar uma conexão com o R5, o que é feito com o comando `setup_r5()`:

```
options(java.parameters = "-Xmx4G")

library(r5r)

r5r_core_before <- setup_r5(dir_before, use_elevation = TRUE, verbose = FALSE)
r5r_core_after <- setup_r5(dir_after, use_elevation = TRUE, verbose = FALSE)
```

Tendo sido criada a conexão com o R5 para os dois cenários, prosseguimos agora para o cálculo da matriz, de fato. De forma a avaliar a expansão do metrô adequadamente, nós precisamos usar os mesmos parâmetros em ambos os cenários. Vamos considerar viagens a pé ou por transporte público, permitir até 800 metros nas pernas de acesso e egresso a paradas e limitar o tempo máximo de viagem em até 60 minutos. Vamos considerar o horário de partida de 07:00 em uma típica segunda-feira de operação. Como origens e destinos, vamos usar os centróides da grade de Fortaleza disponibilizada pelo pacote {aopdata} (mais detalhes sobre o pacote são apresentados na [Seção 5](#)):

```
library(data.table)
library(aopdata)
library(sf)

for_grid <- read_grid("for")
points <- st_centroid(for_grid)

setDT(points)
points <- points[, .(id = id_hex, geom)]
points[, setdiff(names(points), c("id", "geom")) := NULL]
points <- st_sf(points)

ttm_before <- travel_time_matrix(
  r5r_core_before,
  origins = points,
  destinations = points,
  mode = c("WALK", "TRANSIT"),
  departure_datetime = as.POSIXct(
    "02-03-2020 07:00:00",
```



```

    format = "%d-%m-%Y %H:%M:%S"
  ),
  max_walk_dist = 800,
  max_trip_duration = 60,
  verbose = FALSE,
  progress = FALSE
)

ttm_after <- travel_time_matrix(
  r5r_core_after,
  origins = points,
  destinations = points,
  mode = c("WALK", "TRANSIT"),
  departure_datetime = as.POSIXct(
    "02-03-2020 07:00:00",
    format = "%d-%m-%Y %H:%M:%S"
  ),
  max_walk_dist = 800,
  max_trip_duration = 60,
  verbose = FALSE,
  progress = FALSE
)

ttm_before

```

	fromId	toId	travel_time
1:	89801040323ffff	89801040323ffff	2
2:	89801040323ffff	89801040327ffff	22
3:	89801040323ffff	8980104032fffff	14
4:	89801040323ffff	89801040333ffff	9
5:	89801040323ffff	89801040337ffff	30

1665468:	8980107b6dbffff	8980107b6cbffff	8
1665469:	8980107b6dbffff	8980107b6cfffff	14
1665470:	8980107b6dbffff	8980107b6d3ffff	9
1665471:	8980107b6dbffff	8980107b6d7ffff	15
1665472:	8980107b6dbffff	8980107b6dbffff	0

```
ttm_after
```

	fromId	toId	travel_time
--	--------	------	-------------

1:	89801040323ffff	89801040323ffff	2
2:	89801040323ffff	89801040327ffff	22
3:	89801040323ffff	8980104032ffff	14
4:	89801040323ffff	89801040333ffff	9
5:	89801040323ffff	89801040337ffff	30

1671492:	8980107b6dbffff	8980107b6cbffff	8
1671493:	8980107b6dbffff	8980107b6cffff	14
1671494:	8980107b6dbffff	8980107b6d3ffff	9
1671495:	8980107b6dbffff	8980107b6d7ffff	15
1671496:	8980107b6dbffff	8980107b6dbffff	0

À primeira vista, nossas matrizes parecem iguais: todos os tempos de viagem na amostra de pares mostrados acima são idênticos. Isto ocorre porque muitos deslocamentos entre regiões da cidade de fato não são afetados pela implementação da nova linha de metrô, por serem realizados por outros modos de transporte público ou a pé. Se compararmos par a par, no entanto, veremos que vários dos tempos de viagens entre eles são impactados pela Linha Leste:

```
comparison <- merge(
  ttm_before,
  ttm_after,
  by = c("fromId", "toId"),
  suffixes = c("_before", "_after")
)

comparison[travel_time_after < travel_time_before]
```

	fromId	toId	travel_time_before	travel_time_after
1:	89801041b2ffff	8980104c14bffff	55	54
2:	89801041b2ffff	8980104c15bffff	55	53
3:	89801041b2ffff	8980104c397ffff	52	48
4:	8980104532bffff	8980104c13bffff	57	53
5:	89801045363ffff	8980104c13bffff	57	53

8283:	8980107b6dbffff	8980104c02ffff	58	52
8284:	8980107b6dbffff	8980104c13bffff	48	42
8285:	8980107b6dbffff	8980104c14bffff	57	54
8286:	8980107b6dbffff	8980104c15bffff	58	53
8287:	8980107b6dbffff	8980104c397ffff	58	48

6.3 Cálculo da acessibilidade nos cenários antes e depois

O cálculo da acessibilidade é muito simples, requerendo apenas um processamento básico dos nossos dados e a aplicação de uma das funções de acessibilidade do pacote `{accessibility}`. Como tratamento dos dados, vamos juntar as matrizes dos dois cenários em uma única tabela, identificando cada cenário com a coluna `scenario`:

```
ttm <- rbind(ttm_before, ttm_after, idcol = "scenario")
ttm[, scenario := factor(scenario, labels = c("before", "after"))]
```

```
ttm
```

	scenario	fromId	toId	travel_time
1:	before	89801040323ffff	89801040323ffff	2
2:	before	89801040323ffff	89801040327ffff	22
3:	before	89801040323ffff	8980104032fffff	14
4:	before	89801040323ffff	89801040333ffff	9
5:	before	89801040323ffff	89801040337ffff	30

3336964:	after	8980107b6dbffff	8980107b6cbffff	8
3336965:	after	8980107b6dbffff	8980107b6cfffff	14
3336966:	after	8980107b6dbffff	8980107b6d3ffff	9
3336967:	after	8980107b6dbffff	8980107b6d7ffff	15
3336968:	after	8980107b6dbffff	8980107b6dbffff	0

Para o cálculo da acessibilidade, vamos precisar de uma tabela com os dados de uso do solo da cidade de Fortaleza. Para isso, nós usamos a função `read_landuse()` do pacote `{aopdata}`:

```
for_data <- read_landuse("for", showProgress = FALSE)
```

Para fins de demonstração, vamos calcular a acessibilidade a postos de trabalho e a escolas públicas na nossa área de estudo. Os dados do total de empregos e escolas em cada hexágono estão listados nas colunas T001 e E001, respectivamente. Vamos renomeá-las para facilitar sua identificação e manter apenas as colunas necessárias na tabela de dados de uso do solo (vamos manter também as colunas P001, de população total em cada hexágono, e R003, do decil de renda em que cada hexágono se encontra, que serão úteis mais à frente):

```
cols_to_keep <- c("id", "jobs", "schools", "population", "decile")
```

```
setnames(
  for_data,
```

```

    old = c("id_hex", "T001", "E001", "P001", "R003"),
    new = cols_to_keep
)
for_data[, setdiff(names(for_data), cols_to_keep) := NULL]

for_data

```

	id	population	decile	jobs	schools
1:	89801040323ffff	30	1	0	0
2:	89801040327ffff	318	1	7	0
3:	8980104032bffff	0	NA	0	0
4:	8980104032fffff	103	1	98	0
5:	89801040333ffff	43	1	0	0

2558:	8980107b6cbffff	2575	4	124	0
2559:	8980107b6cfffff	2997	3	4	0
2560:	8980107b6d3ffff	1751	8	14	0
2561:	8980107b6d7ffff	2032	4	134	0
2562:	8980107b6dbffff	1896	9	193	1

A parte mais difícil do processo do cálculo de acessibilidade é a definição da medida a ser utilizada. É extremamente importante pesar as vantagens e desvantagens de cada uma e compreender quais medidas se adequam às oportunidades para quais se deseja calcular os níveis de acessibilidade. Neste exemplo, utilizaremos duas medidas distintas:

- No cálculo da acessibilidade a empregos, vamos usar a medida de oportunidades cumulativas. Esta métrica nos permite entender quantos empregos são acessíveis dentro de um determinado custo de viagem, representando, desta forma, a oferta de empregos aos habitantes de cada uma das células de nossa grade. Embora essa medida não leve em consideração a competição por empregos, sua principal vantagem é que ela é de muito simples comunicação e interpretação. Em nosso exemplo, vamos tomar como limite de custo um tempo de viagem de 30 minutos.
- No cálculo da acessibilidade a escolas públicas, vamos usar a medida de custo mínimo de viagem. Esta métrica é especialmente útil para avaliar a cobertura de serviços públicos essenciais, como educação e saúde básica, pois é desejado que a maior parte da população de uma cidade esteja a uma distância considerada razoável desses serviços.

Como mostrado anteriormente, na [Section 3.2](#), o cálculo dessas medidas pode ser feito com as funções `cumulative_cutoff()` e `cost_to_closest()`, respectivamente, do pacote `{accessibility}`:

```

library(accessibility)

setnames(ttm, old = c("fromId", "toId"), new = c("from_id", "to_id"))

access_to_jobs <- cumulative_cutoff(
  travel_matrix = ttm,
  land_use_data = for_data,
  opportunity = "jobs",
  travel_cost = "travel_time",
  cutoff = 60,
  group_by = "scenario"
)
access_to_jobs

```

	id	scenario	jobs
1:	89801040323ffff	before	46159
2:	89801040327ffff	before	25499
3:	8980104032bffff	before	118
4:	8980104032fffff	before	66513
5:	89801040333ffff	before	46159

5120:	8980107b6cbffff	after	366197
5121:	8980107b6cfffff	after	256888
5122:	8980107b6d3ffff	after	324981
5123:	8980107b6d7ffff	after	338965
5124:	8980107b6dbffff	after	361967

```

time_to_schools <- cost_to_closest(
  travel_matrix = ttm,
  land_use_data = for_data,
  opportunity = "schools",
  travel_cost = "travel_time",
  group_by = "scenario"
)
time_to_schools

```

	id	scenario	travel_time
1:	89801040323ffff	before	25
2:	89801040323ffff	after	25
3:	89801040327ffff	before	35

4:	89801040327ffff	after	35
5:	8980104032bffff	before	Inf

5120:	8980107b6d3ffff	after	9
5121:	8980107b6d7ffff	before	13
5122:	8980107b6d7ffff	after	13
5123:	8980107b6dbffff	before	0
5124:	8980107b6dbffff	after	0

Como podemos ver, o resultado da função de custo mínimo de viagem inclui alguns valores `Inf`. As origens listadas com este valor não conseguem alcançar dentro de 120 minutos (limite de tempo de viagem imposto no cálculo da matriz) nenhum destino em que está localizada ao menos uma escola pública. O valor `Inf`, portanto, é utilizado para sinalizar esta condição.

6.4 Análise dos níveis de acessibilidade antes e depois

Agora que calculamos os níveis de acessibilidade em cada cenário, precisamos compreender como as mudanças na frequência das linhas de transporte público afetaram as condições de acessibilidade em nossa área de estudo. Uma boa forma de entender isso é olhando para a distribuição espacial dos níveis de acessibilidade. Vamos juntar usar as informações espaciais da nossa área de estudo, disponíveis no objeto `for_grid`, baixado do pacote `{aopdata}` anteriormente, às nossas tabelas de acessibilidade:

```
setDT(for_grid)

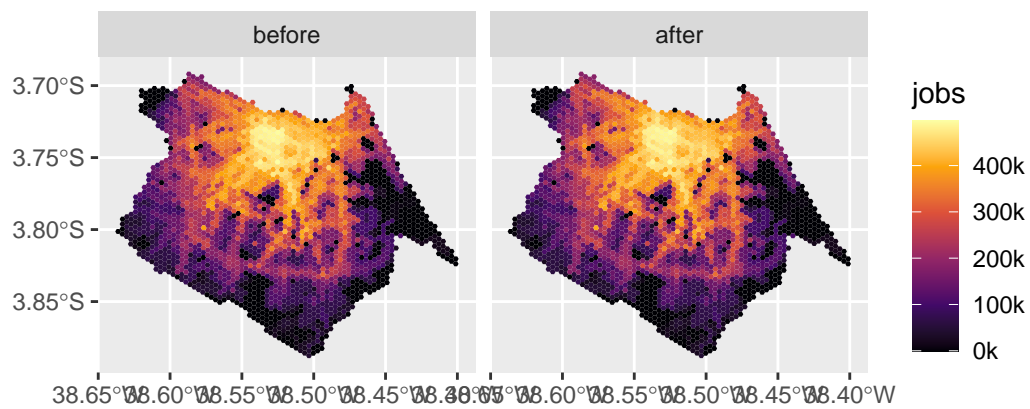
access_to_jobs[for_grid, on = c(id = "id_hex"), geom := i.geom]
time_to_schools[for_grid, on = c(id = "id_hex"), geom := i.geom]
```

Feito isso, apresentamos os resultados em formato de mapa:

```
library(ggplot2)

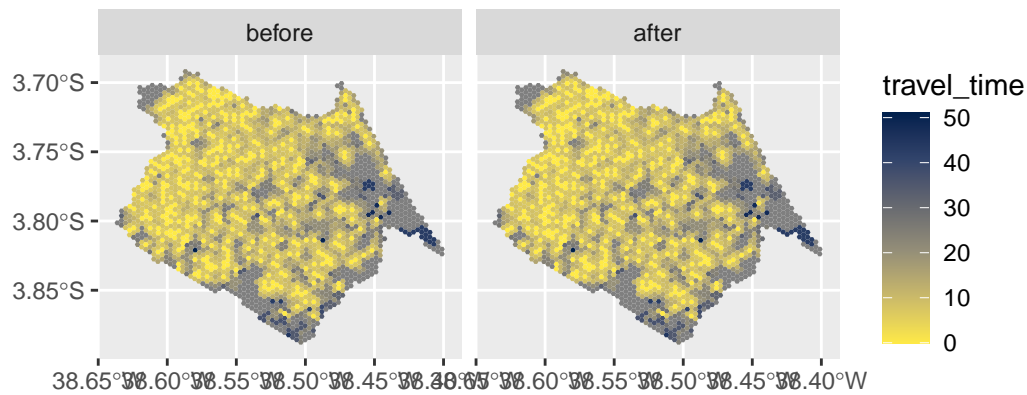
access_to_jobs_sf <- st_sf(access_to_jobs)

ggplot(access_to_jobs_sf) +
  geom_sf(aes(fill = jobs), color = NA) +
  facet_wrap(~ scenario, nrow = 1) +
  scale_fill_viridis_c(
    option = "inferno",
    label = scales::label_number(scale = 1/1000, suffix = "k")
  )
```



```
time_to_schools_sf <- st_sf(time_to_schools)

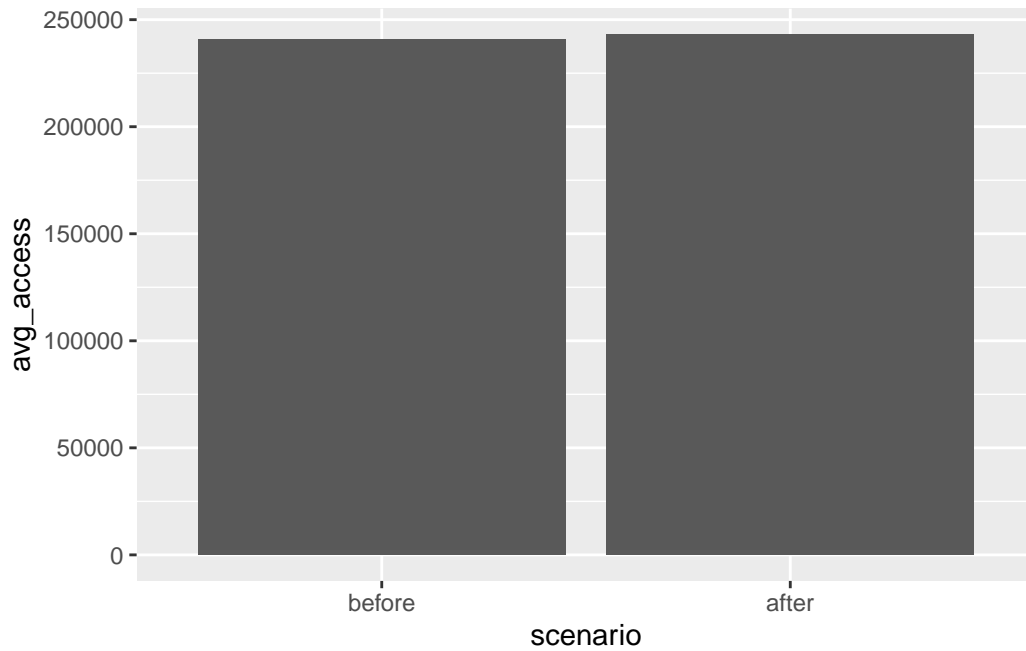
ggplot(time_to_schools_sf) +
  geom_sf(aes(fill = travel_time), color = NA) +
  facet_wrap(~ scenario, nrow = 1) +
  scale_fill_viridis_c(option = "cividis", direction = -1)
```



Como podemos ver, a construção da Linha Leste do metrô de Fortaleza por si só, ignorando as demais intervenções que acompanham esta intervenção, pouco afeta os níveis de acessibilidade na cidade. Isto fica claro quando olhamos para os níveis médios de acessibilidade nos dois cenários, que pouco se alteram:

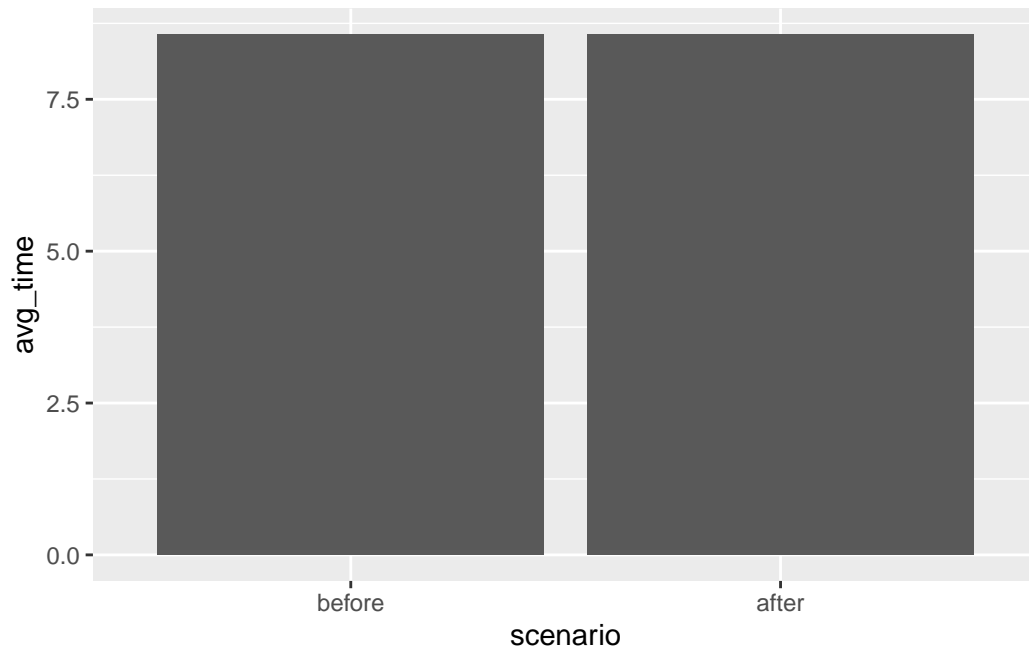
```
access_to_jobs[for_data, on = "id", population := i.population]
avg_access_to_jobs <- access_to_jobs[
  ,
  .(avg_access = weighted.mean(jobs, w = as.numeric(population))),
  by = scenario
]

ggplot(avg_access_to_jobs) + geom_col(aes(scenario, avg_access))
```

```
time_to_schools[for_data, on = "id", population := i.population]
time_to_schools[
  ,
  treated_travel_time := ifelse(is.infinite(travel_time), 120, travel_time)
]
avg_time_to_schools <- time_to_schools[
  ,
  .(avg_time = weighted.mean(treated_travel_time, w = population)),
  by = scenario
]

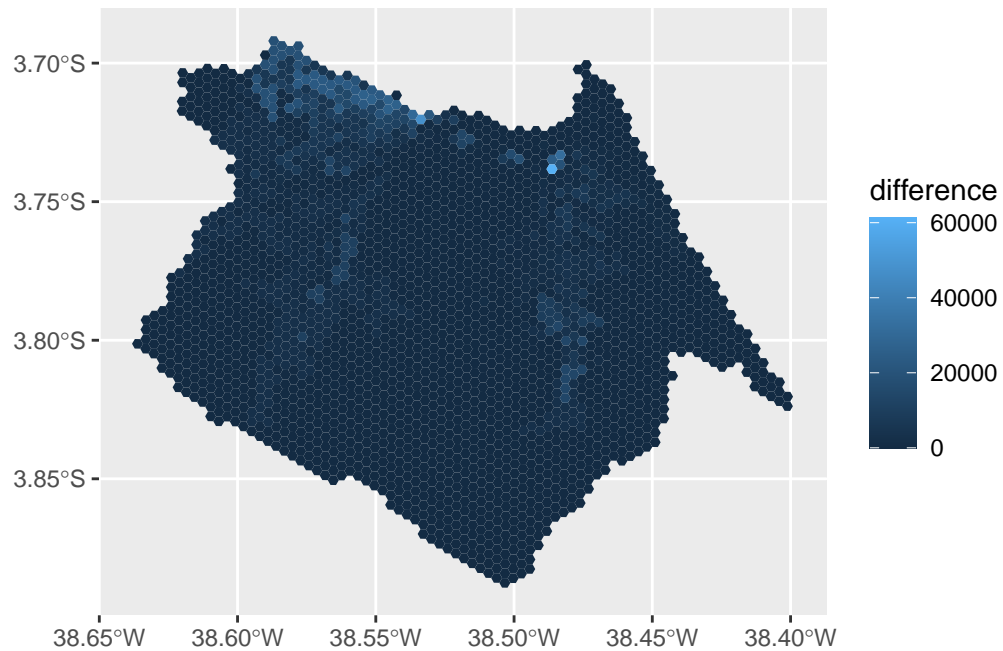
ggplot(avg_time_to_schools) + geom_col(aes(scenario, avg_time))
```



Neste caso, uma forma de mostrar o impacto da política de forma mais clara é apresentar a distribuição espacial da diferença de acessibilidade entre os dois cenários:

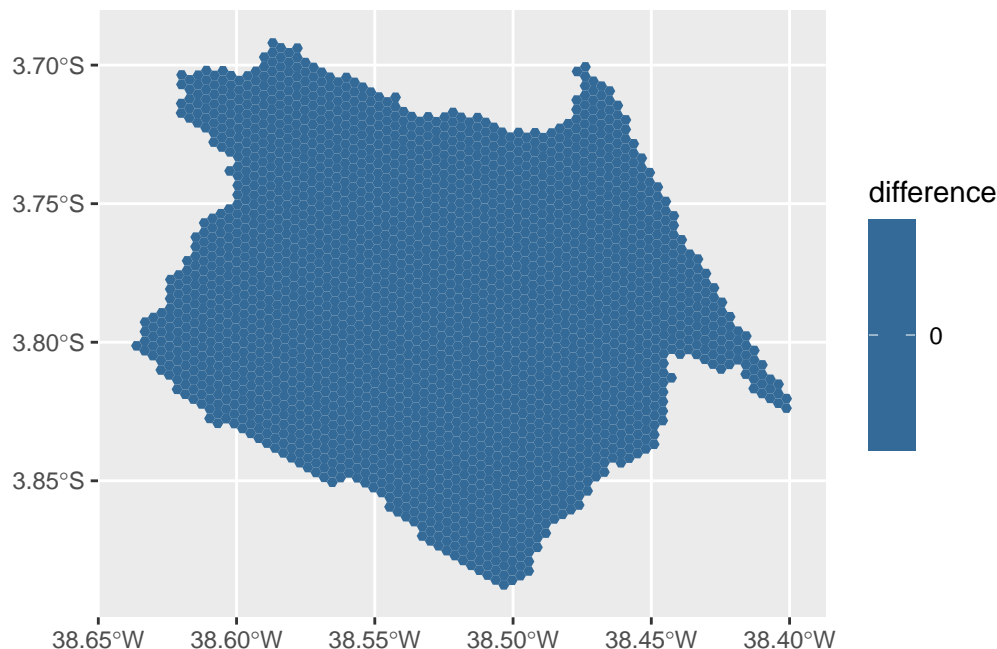
```
access_to_jobs_diff <- dcast(
  access_to_jobs,
  id ~ scenario,
  value.var = "jobs"
)
access_to_jobs_diff[, difference := after - before]
access_to_jobs_diff[for_grid, on = c(id = "id_hex"), geom := i.geom]

ggplot(st_sf(access_to_jobs_diff)) +
  geom_sf(aes(fill = difference), color = NA)
```



```
time_to_schools_diff <- dcast(
  time_to_schools,
  id ~ scenario,
  value.var = "travel_time"
)
time_to_schools_diff[, difference := after - before]
time_to_schools_diff[is.nan(difference), difference := 0]
time_to_schools_diff[for_grid, on = c(id = "id_hex"), geom := i.geom]

ggplot(st_sf(time_to_schools_diff)) +
  geom_sf(aes(fill = difference), color = NA)
```



Os ganhos de acessibilidade ao emprego se concentram principalmente próximos às estações do novo corredor e dos corredores já existentes que se ligam a ele. Isto mostra como a implantação de uma infraestrutura de transporte pode impactar não apenas suas imediações, mas também locais mais afastados, graças à conectividade da rede.

Por sua vez, o mapa de diferença no tempo de acesso a escolas apresenta um cenário diferente: nenhuma região da cidade apresenta melhorias no tempo de acesso. Diferentemente da distribuição de empregos, que tende a seguir critérios econômicos e mercadológicos, a distribuição de equipamentos públicos, entre eles escolas públicas, tende a ser planejada usando critérios de justiça social e equidade, sendo, portanto, mais equilibrada no território. Dessa forma, a construção da Linha Leste na parte central da cidade em nada afeta o tempo de acesso das pessoas que moram em locais afastados desse corredor. Não apenas isso, mas mesmo para as pessoas que moram perto do novo corredor a sua implementação não muda os níveis de acesso a escolas. Isto é um sinal que esta área possui uma boa oferta de escolas e já é bem servida de transporte público, mesmo antes da finalização da Linha Leste.

6.5 Desigualdade de acessibilidade

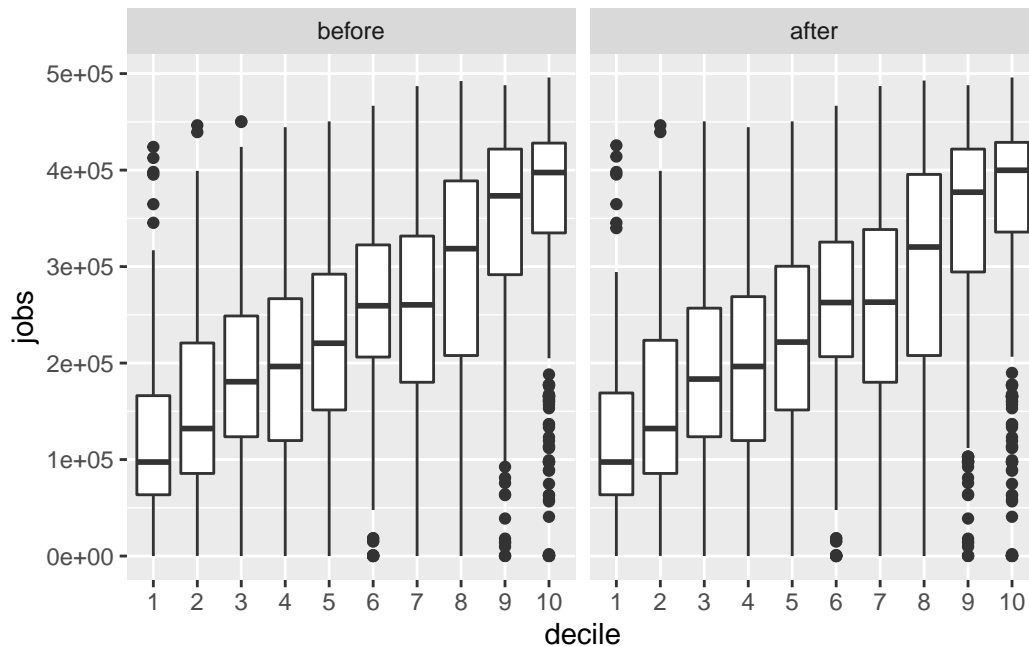
Políticas que resultam em maiores ganhos de acessibilidade para determinados grupos populacionais em detrimento de outros, como é o caso observado nos níveis de acesso a empregos em Fortaleza, alteram também a distribuição da acessibilidade entre grupos socioeconômicos na

cidade. Em outras palavras, políticas que trazem mais benefícios para certas pessoas do que para outras mudam a configuração da desigualdade de acessibilidade.

Nós podemos entender como é a distribuição de níveis de acessibilidade entre grupos socioeconômicos usando um gráfico *box plot* e agrupando a nossa população por decis de renda. Separando a nossa visualização nos cenários antes e depois, compreendemos como a Linha Leste afeta os níveis de acessibilidade de decil de renda:

```
for_data[, decile := factor(decile)]
access_to_jobs[for_data, on = "id", decile := i.decile]

ggplot(access_to_jobs[!is.na(decile)]) +
  geom_boxplot(aes(decile, jobs, weight = population, group = decile)) +
  facet_wrap(~ scenario, nrow = 1)
```

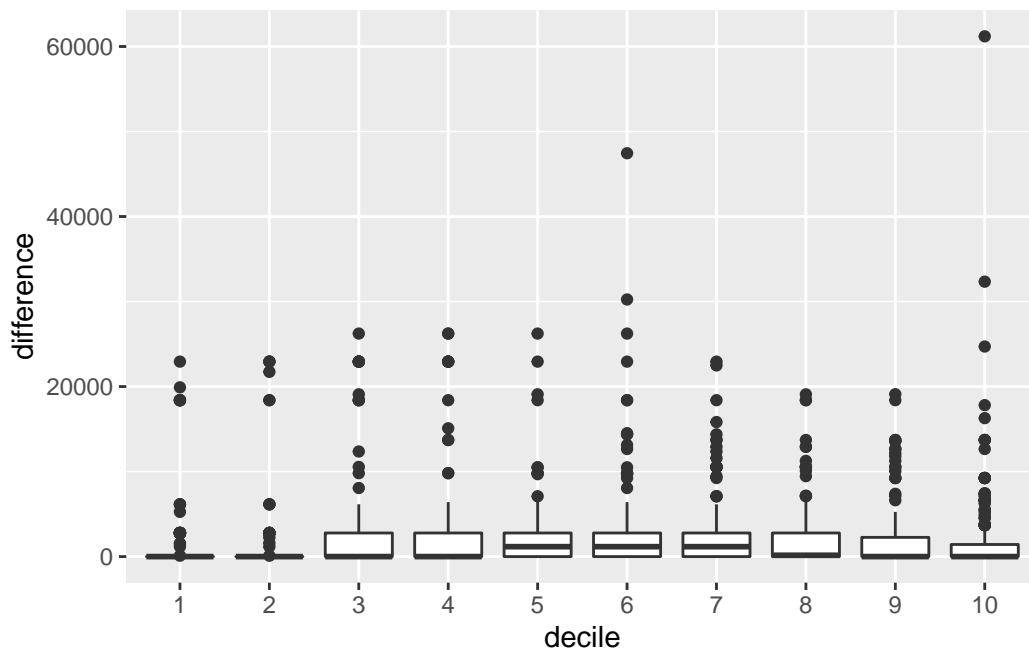


Os dois painéis mostram claramente que em Fortaleza as pessoas mais ricas detêm os maiores níveis de acessibilidade ao emprego. Em Fortaleza, assim como na maior parte das grandes cidades brasileiras, os mais ricos tendem a morar próximos às grandes concentrações de emprego da cidade, enquanto os mais pobres tendem a ocupar regiões periféricas. Consequentemente, os níveis de acessibilidade dos mais ricos costuma ser maiores do que os dos mais pobres - não apenas porque moram mais perto dos empregos, mas porque também, geralmente, regiões centrais costumam ser mais bem servidas de transporte público do que as periferias.

Assim como no caso da distribuição espacial da acessibilidade, é muito difícil enxergar diferenças consideráveis nos níveis de acesso entre os cenários antes e depois. Lançando mão do mesmo recurso que fizemos anteriormente, podemos enxergar essas diferenças mais claramente visualizando a distribuição da diferença de acessibilidade entre cada cenário por decil de renda:

```
access_to_jobs_diff[
  for_data,
  on = "id",
  `:=`(
    decile = i.decile,
    population = i.population
  )
]

ggplot(access_to_jobs_diff[!is.na(decile)]) +
  geom_boxplot(aes(decile, difference, weight = population, group = decile))
```



Como podemos ver, os ganhos de acessibilidade se concentram principalmente entre as camadas médias da população, em especial os decis 5, 6 e 7. No entanto, o hexágono que mostrou maior aumento de acessibilidade aparece como *outlier* do 10º decil, apresentando um ganho de mais de 60000 empregos em relação ao cenário pré-implementação.

Uma forma frequentemente empregada de resumir o panorama da desigualdade de uma determinado local é com o uso de uma medida de desigualdade. Esse tipo de medida busca resumir a distribuição dos níveis de acessibilidade entre grupos populacionais (aqui, decis de renda) em um único indicador que facilite a comunicação dos resultados e que possa ser usado, por exemplo, na elaboração de planos e metas. As medidas de desigualdade mais frequentemente utilizadas na literatura de acessibilidade são a Razão de Palma e o Índice de Gini.

Neste exemplo, vamos calcular a Razão de Palma dos cenários antes e depois. Esta medida é o resultado da divisão da acessibilidade média dos 10% mais ricos pela acessibilidade média dos 40% mais pobres:

$$P = \frac{\overline{A_{tp10}}}{\overline{A_{bt40}}}$$

em que P é a Razão de Palma, $\overline{A_{tp10}}$ é a acessibilidade média dos 10% mais ricos e $\overline{A_{bt40}}$ é a acessibilidade média dos 40% mais pobres.

A principal vantagem da Razão de Palma frente ao Índice de Gini é a sua facilidade de comunicação e interpretação. Valores maiores do que 1 indicam um cenário em que os ricos possuem níveis médios do que os pobres, e valores menores do que 1 a situação inversa. Outra vantagem da Razão de Palma é que ela reflete claramente como a desigualdade varia entre dois grupos de especial interesse para nós: os mais privilegiados e os mais vulneráveis de uma população. O Índice de Gini, por sua vez, estima o quanto uma distribuição desvia de uma distribuição hipotética em que todos possuem o mesmo nível de acesso, mas nada diz sobre as condições socioeconômicas daqueles que detêm os maiores e menores níveis de acessibilidade. Se uma determinada política aumenta os níveis de acessibilidades de pessoas de alta renda que moram em locais de baixa acessibilidade, por exemplo, o Índice de Gini indicaria que houve uma diminuição da desigualdade, mesmo que nenhuma população vulnerável se beneficie dessa política. Esta dificilmente é, entretanto, uma política equitativa. Neste mesmo caso, a Razão de Palma indicaria um aumento da desigualdade, resultado mais alinhado às expectativas.

Analisar a Razão de Palma nos cenários antes e depois, dessa forma, nos permite entender qual o impacto da Linha Leste na desigualdade de acesso a empregos em Fortaleza:

```
richest_access <- access_to_jobs[
  decile == 10,
  .(access = weighted.mean(jobs, w = population)),
  by = scenario
]

poorest_access <- access_to_jobs[
  decile %in% 1:4,
  .(access = weighted.mean(jobs, w = as.numeric(population))),
```

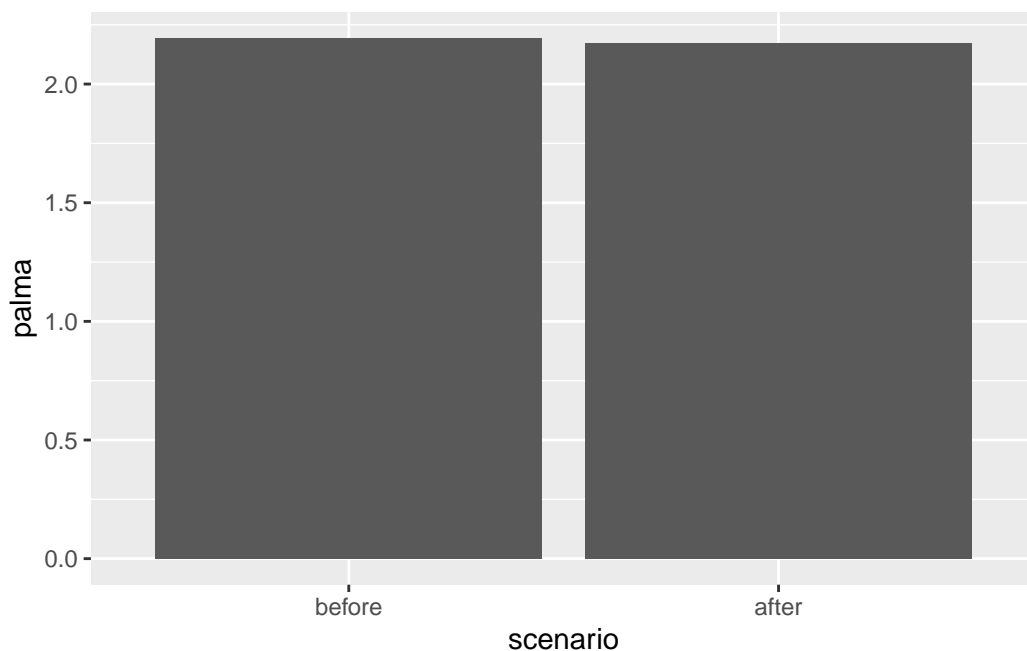
```

    by = scenario
  ]

  palma_ratio <- merge(
    richest_access,
    poorest_access,
    by = "scenario",
    suffixes = c("_richest", "_poorest")
  )
  palma_ratio[, palma := access_richest / access_poorest]

  ggplot(palma_ratio) + geom_col(aes(scenario, palma))

```



Como podemos ver, a Razão de Palma sofreu uma ligeira queda entre os cenários pré e pós implementação da Linha Leste. Consequentemente, podemos concluir que a expansão do metrô de Fortaleza seria uma política progressiva neste nosso cenário simplificado, que leva em consideração apenas a construção do novo corredor e ignora ajustes nas frequências de outros corredores e mudanças nas linhas de ônibus da cidade.

Part V

SEÇÃO 5: Dados do Projeto AOP

Objetivo: O objetivo deste capítulo é mostrar como fazer download e analisar os dados do projeto Acesso a Oportunidades (AOP) utilizando o pacote `{aopdata}` no R.

Nos capítulos anteriores você aprendeu como calcular indicadores de acessibilidade. No entanto, em muitos casos, você não tem disponibilidade para calcular esses indicadores por conta própria e tudo o que você quer é analisar os resultados que já foram calculados por alguém.

O projeto [Acesso a Oportunidades](#) disponibiliza uma extensa base de dados com informações sobre a distribuição da população, atividades econômicas e serviços públicos, além de várias estimativas de acessibilidade urbana para diversos tipos de atividades. Essas estimativas de acessibilidade estão disponíveis para diferentes modos de transporte (caminhada, bicicleta, transporte público e automóvel), horários do dia (pico e fora-pico), grupos populacionais (segundo níveis de renda, cor, sexo e idade) e para diferentes atividades (empregos, escolas, serviços de saúde e centros de assistência social). Nesta versão, a base de dados traz essas informações para diversos anos (2017, 2018 e 2019), se apoiando em uma única metodologia consistente para as 20 maiores cidades do Brasil. Veja abaixo Table 6.1 e Table 6.2.

i As metodologias utilizadas para gerar estes dados são apresentadas em detalhe em publicações separadas, para os dados populacionais e de uso do solo (Rafael H. M. Pereira et al. 2022), e para os dados de acessibilidade ([pereira2022estimativas?](#)).

Quais dados estão disponíveis?

Escopo dos dados



Figure 6.1: Escopo dos dados do AOP

Dados de população, empregos e serviços públicos

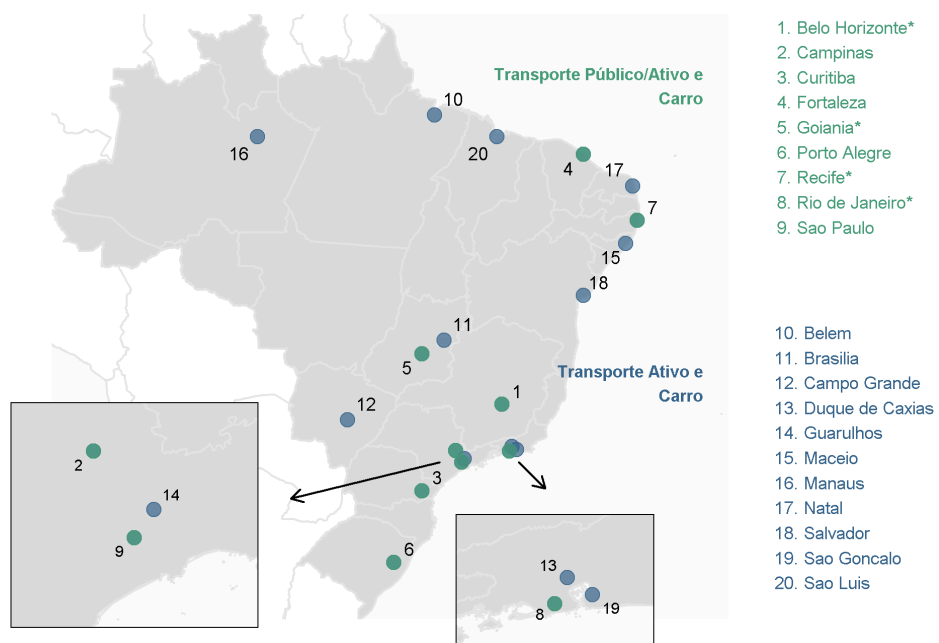


Figure 6.2: Escopo geográfico dos dados do AOP

Table 6.1: Informações socioeconômicas da população e de distribuição espacial de atividades, segundo ano e fonte de dados

Dado	Informações	Anos	Fonte
Características sociodemográficas da população	Quantidade de pessoas segundo sexo, faixa de idade e cor/raça; média da renda domiciliar per capita	2010	Censo Demográfico, IBGE
Estabelecimentos de educação	Quantidade de creches e escolas públicas segundo nível infantil, fundamental e médio	2017, 2018, 2019	Censo Escolar, Inep
Estabelecimentos de saúde	Quantidade de estabelecimentos de saúde que atendem pelo SUS segundo nível de atenção baixa, média e alta complexidade	2017, 2018, 2019	Cadastro Nacional de Estabelecimentos de Saúde (CNES), Ministério da Saúde
Atividade econômica	Quantidade de empregos formais conforme o nível de instrução dos trabalhadores baixa, média e alta escolaridade	2017, 2018, 2019	Relação Anual de Informações Sociais (RAIS), Ministério da Economia
Estabelecimentos de assistência social	Quantidade de CRAS	2017, 2018, 2019	Censo SUAS, Ministério da Cidadania

Dados de acessibilidade urbana

Table 6.2: Indicadores de acessibilidade calculados no Projeto Acesso a Oportunidades

Indicador (código)	Descrição	Tipo de oportunidades	Limites de tempo de viagem
Tempo mínimo de viagem (TMI)	Tempo até a oportunidade mais próxima	Saúde, Educação, CRAS	A pé (60 mins); Bicicleta, Transporte público e carro (120 mins)
Medida cumulativa ativa (CMA)	Quantidade de oportunidades acessíveis em um determinado limite de tempo	Trabalho, Saúde, Educação, CRAS	A pé e bicicleta (15, 30, 45 e 60 mins); Transporte público e carro (15, 30, 60, 90 e 120 mins)

Indicador (código)	Descrição	Tipo de oportu- nidades	Limites de tempo de viagem
Medida cumulativa passiva (CMP)	Quantidade de pessoas que acessam a localidade em um determinado limite de tempo	-	A pé e bicicleta (15, 30, 45 e 60 mins); Transporte público e carro (15, 30, 60, 90 e 120 mins)

Todas as bases de dados criadas pelo Projeto Acesso a Oportunidades (AOP) estão disponíveis para download no [site do projeto](#) ou pelo pacote de R `{aopdata}`. Nas próximas seções são apresentados exemplos de como baixar e visualizar esses dados em R.

7 Dados de população e socioeconômicos

7.1 Download dos dados

Para fazer o download dos dados do projeto AOP usando o pacote `aopdata`, você pode usar a função `read_population()`. Essa função baixa estimativas do Censo Demográfico de 2010 [ref] do IBGE sobre a distribuição espacial da população e suas características em termos de renda domiciliar per capita, cor, sexo e idade. Nesta função, o parâmetro `city` permite você indicar os dados de qual cidade serão baixados.

Os dados estão agregados espacialmente em uma grade de [hexágonos H3](#) na resolução 9, na qual cada hexágono tem uma área de 0.11 km², o que equivale a aproximadamente o tamanho de um quarteirão. Para baixar os dados com as informações espaciais de geometria da grade espacial, você deve usar o parâmetro `geometry = TRUE`.

Neste exemplo, abaixo, nós mostramos como baixar os dados de população do Censo de 2010 para Fortaleza.

```
library(aopdata)

# download aop population data
df <- read_population(
  city = "Fortaleza",
  year = 2010,
  geometry = TRUE,
  showProgress = FALSE
)
```

Os dados da tabela tem essa aparência aqui:

```
head(df)
```

```
Simple feature collection with 6 features and 22 fields
Geometry type: POLYGON
Dimension:      XY
Bounding box:   xmin: -38.50828 ymin: -3.889301 xmax: -38.4983 ymax: -3.878958
Geodetic CRS:   WGS 84
```

	year	id_hex	abbrev_muni	name_muni	code_muni	P001	P002	P003	P004	P005
1	2010	89801040323fffff		for Fortaleza	2304400	30	8	21	0	1
2	2010	89801040327fffff		for Fortaleza	2304400	318	77	233	0	8
3	2010	8980104032bfffff		for Fortaleza	2304400	0	0	0	0	0
4	2010	8980104032fffff		for Fortaleza	2304400	103	24	77	0	2
5	2010	89801040333fffff		for Fortaleza	2304400	43	11	31	0	1
6	2010	89801040337fffff		for Fortaleza	2304400	348	86	252	0	10

	P006	P007	P010	P011	P012	P013	P014	P015	P016	R001	R002	R003
1	17	13	3	4	2	3	8	9	1	168.6	1	1
2	168	150	30	50	26	38	80	82	12	202.6	1	1
3	0	0	0	0	0	0	0	0	0	NA	NA	NA
4	53	50	10	16	8	13	25	27	4	245.6	1	1
5	22	21	4	7	3	5	11	12	1	187.3	1	1
6	175	173	34	54	27	41	89	89	14	168.6	1	1


```

geometry
1 POLYGON ((-38.50232 -3.8858...
2 POLYGON ((-38.50527 -3.8840...
3 POLYGON ((-38.49932 -3.8841...
4 POLYGON ((-38.50227 -3.8824...
5 POLYGON ((-38.50237 -3.8893...
6 POLYGON ((-38.50532 -3.8875...

```

De imediato, se nota que os nomes das variáveis (colunas) da base de dados estão organizadas com códigos, como P001, P002, ..., R001, R002, etc. A descrição completa do dicionário de variáveis está disponível [aqui](#). A descrição de algumas dessas colunas é apresentada nas próximas seções, onde mostramos como fazer a visualização de alguns desses dados em mapas e gráficos.

7.2 Mapa de população total

Antes de visualizar os dados do {aopdata}, nós precisamos carregar algumas bibliotecas de visualização e manipulação de dados.

```

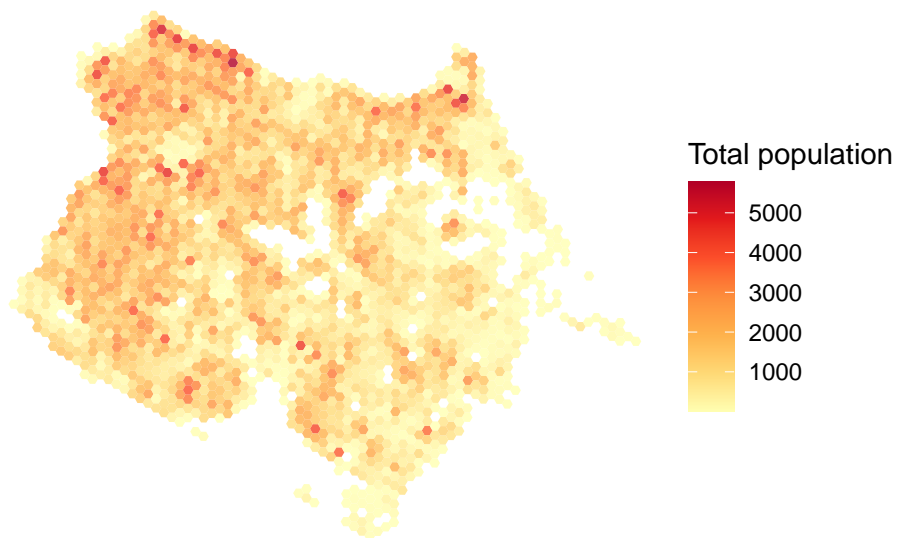
library(patchwork)
library(ggplot2)
library(scales)
library(sf)

```

Com um comando, é possível visualizar a distribuição espacial da população de Fortaleza. A figura mostra um mapa coroplético onde a cor de cada célula da grade espacial é preenchida com base na quantidade total de pessoas residentes (variável P001).

```
ggplot(subset(df, P001 > 0)) +
  geom_sf(aes(fill = P001), color = NA, alpha = 0.8) +
  scale_fill_distiller(palette = "YlOrRd", direction = 1) +
  labs(
    title = "Population distribution",
    subtitle = "Fortaleza",
    fill = "Total population"
  ) +
  theme_void()
```

Population distribution
Fortaleza



7.3 Mapa de população por cor

Além da informação sobre a população total em cada célula, o dados do `{aopdata}` também permitem saber a quantidade de pessoas de diferentes cores (variáveis P002 a P005), sexo (variáveis P006 e P007) e faixas etárias (variáveis P010 a P016). O código abaixo ilustra como é simples calcular a proporção de pessoas negras e brancas em cada hexágono e visualizar esses dados num mapa.

```
pop_b <- ggplot(subset(df, P001 > 0)) +
  geom_sf(aes(fill = P003 / P001), color = NA, alpha = 0.8) +
```



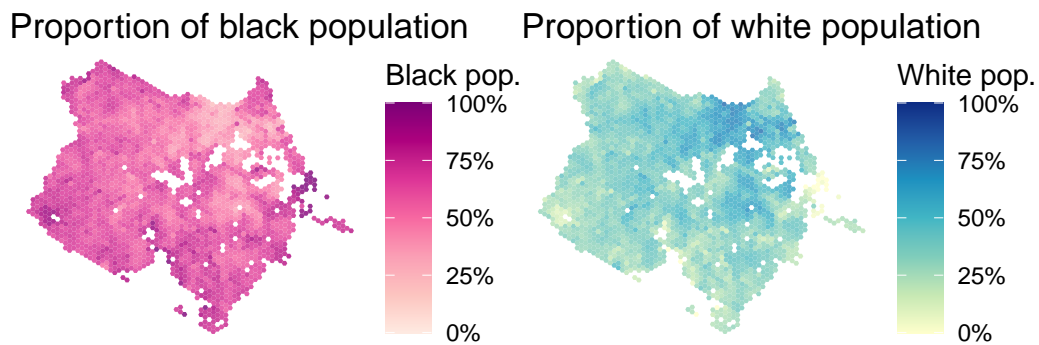
```

scale_fill_distiller(
  palette = "RdPu",
  direction = 1,
  labels = percent,
  limits = c(0, 1)
) +
labs(title = "Proportion of black population", fill = "Black pop.") +
theme_void()

pop_w <- ggplot(subset(df, P001 > 0)) +
  geom_sf(aes(fill = P002 / P001), color = NA, alpha = 0.8) +
  scale_fill_distiller(
    palette = "YlGnBu",
    direction = 1,
    labels = percent,
    limits = c(0, 1)
  ) +
  labs(title = "Proportion of white population", fill = "White pop.") +
  theme_void()

pop_b + pop_w

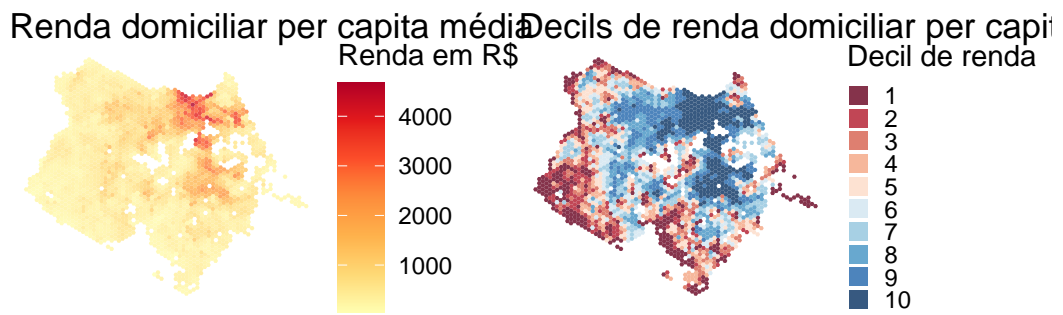
```



7.4 Mapa de população por renda

Os dados trazem também informação sobre a renda domiciliar per capita média de cada hexágono (R001), e sua classificação em termos de quintil (R002) e decil de renda (R003). Com esses dados, é possível visualizar com o comando abaixo a distribuição espacial dos diferentes níveis de renda da cidade.R

```
renda_c <- ggplot(subset(df, P001 > 0)) +  
  geom_sf(aes(fill = R001), color = NA, alpha = 0.8) +  
  scale_fill_distiller(palette = "YlOrRd", direction = 1) +  
  labs(title = "Renda domiciliar per capita média", fill = "Renda em R$") +  
  theme_void()  
  
renda_d <- ggplot(subset(df, !is.na(R002))) +  
  geom_sf(aes(fill = factor(R003)), color = NA, alpha = 0.8) +  
  scale_fill_brewer(palette = "RdBu") +  
  labs(  
    title = "Decils de renda domiciliar per capita",  
    fill = "Decil de renda"  
  ) +  
  theme_void() +  
  theme(legend.key.size = unit(0.3, "cm"))  
  
renda_c + renda_d
```



8 Dados de distribuição espacial de oportunidades

8.1 Download dos dados

O pacote `{aopdata}` também permite baixar, para todas cidades incluídas no projeto, dados anuais da distribuição espacial de empregos (de baixa, média e alta escolaridade), estabelecimentos de saúde (de baixa, média e alta complexidade), escolas pública (ensino infantil, fundamental e médio), e de centros de referência de assistência social (Cras).

Todos esses dados podem ser baixados com a função `read_landuse()`, que funciona da mesma maneira que a função `read_population()`. Você só precisa indicar nos parâmetros da função qual cidade (`city`) e ano (`year`) devem ser baixados, além de apontar se deseja que os dados contenham as informações espaciais dos hexágonos (`geometry = TRUE`).

Neste exemplo, abaixo, nós mostramos como baixar os dados de uso do solo no ano de 2019 para Belo Horizonte. Note que essa função automaticamente já baixa também os dados de população.

```
library(aopdata)

df <- read_landuse(
  city = "Belo Horizonte",
  year = 2019,
  geometry = TRUE,
  showProgress = FALSE
)

head(df)
```

Simple feature collection with 6 features and 39 fields

Geometry type: POLYGON

Dimension: XY

Bounding box: xmin: -43.87914 ymin: -19.86084 xmax: -43.85906 ymax: -19.82421

Geodetic CRS: WGS 84

id_hex	abbrev_muni	name_muni	code_muni	P001	P002	P003	P004	P005
--------	-------------	-----------	-----------	------	------	------	------	------

1	89a881345a3ffff	bho	Belo Horizonte	3106200	0	0	0	0	0						
2	89a881345a7ffff	bho	Belo Horizonte	3106200	0	0	0	0	0						
3	89a881345b7ffff	bho	Belo Horizonte	3106200	0	0	0	0	0						
4	89a88136103ffff	bho	Belo Horizonte	3106200	508	141	353	4	10						
5	89a88136107ffff	bho	Belo Horizonte	3106200	428	114	304	3	7						
6	89a8813610bffff	bho	Belo Horizonte	3106200	399	111	277	3	8						
	P006	P007	P010	P011	P012	P013	P014	P015	P016	R001	R002	R003	year	T001	T002
1	0	0	0	0	0	0	0	0	0	NA	NA	NA	2019	0	0
2	0	0	0	0	0	0	0	0	0	NA	NA	NA	2019	0	0
3	0	0	0	0	0	0	0	0	0	NA	NA	NA	2019	0	0
4	253	255	53	86	30	43	176	110	10	502.9	2	3	2019	100	17
5	210	218	42	72	27	38	142	98	9	491.8	2	3	2019	44	19
6	203	196	42	68	24	34	138	86	7	502.8	2	3	2019	104	31
	T003	T004	E001	E002	E003	E004	M001	M002	M003	M004	S001	S002	S003	S004	C001
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	72	11	1	1	0	0	321	321	0	0	0	0	0	0	0
5	23	2	0	0	0	0	0	0	0	0	0	0	0	0	0
6	58	15	1	0	1	0	817	0	817	0	0	0	0	0	0
	geometry														
1	POLYGON ((-43.86011 -19.829...														
2	POLYGON ((-43.86313 -19.827...														
3	POLYGON ((-43.86321 -19.830...														
4	POLYGON ((-43.8731 -19.8608...														
5	POLYGON ((-43.87612 -19.859...														
6	POLYGON ((-43.87001 -19.859...														

Antes de visualizar os dados de uso do solo nas próximas seções, vamos carregar algumas bibliotecas de visualização e manipulação de dados.

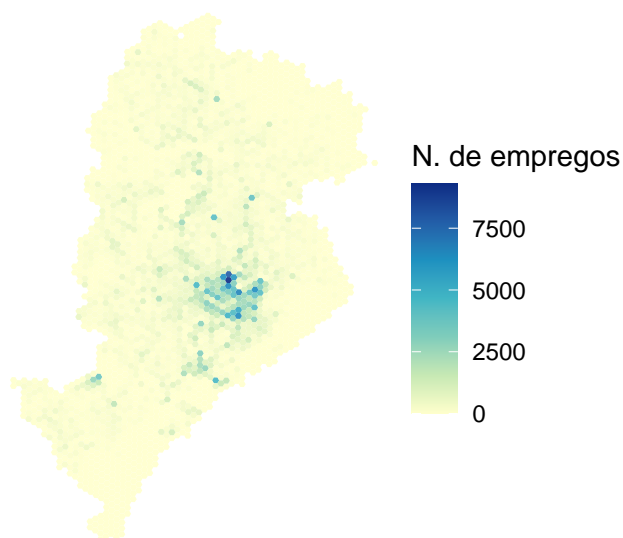
```
library(patchwork)
library(ggplot2)
library(scales)
library(sf)
```

8.2 Mapa de empregos

Como nos exemplos anteriores, é possível visualizar o mapa de distribuição espacial de empregos usando o pacote `ggplot2` com o código abaixo:

```
ggplot(df) +
  geom_sf(aes(fill = T001), color = NA, alpha = 0.9) +
  scale_fill_distiller(palette = "YlGnBu", direction = 1) +
  labs(
    title = "Distribuição espacial de empregos",
    subtitle = "Belo Horizonte",
    fill = "N. de empregos"
  ) +
  theme_void()
```

Distribuição espacial de empregos
Belo Horizonte



8.3 Mapa de escolas

As variáveis que indicam o número de escolas em cada célula são aquelas que começam com a letra E. Neste exemplo abaixo, nós mapeamos a distribuição espacial de todas as escolas de Belo Horizonte a partir da variável E001.

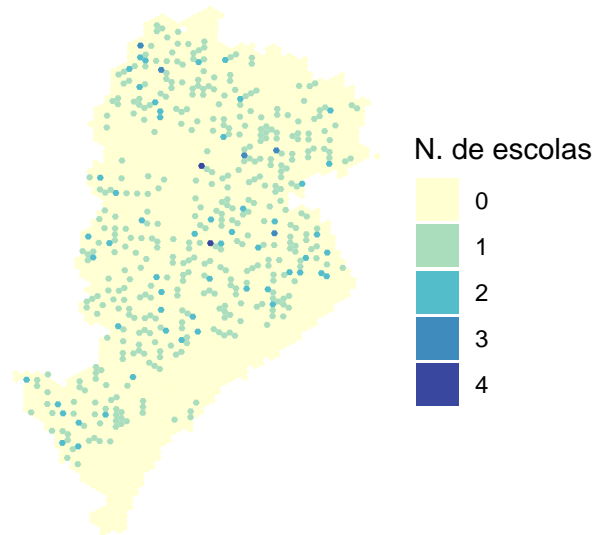
```
ggplot(df) +
  geom_sf(aes(fill = as.factor(E001)), color = NA, alpha = 0.9) +
  scale_fill_brewer(palette = "YlGnBu", direction = 1) +
  labs(
```

```

title = "Distribuição espacial de escolas",
subtitle = "Belo Horizonte",
fill = "N. de escolas"
) +
theme_void()

```

Distribuição espacial de escolas Belo Horizonte



8.4 Mapa de estabelecimentos de saúde

Para analisar a distribuição espacial de estabelecimentos de saúde, você deve analisar as variáveis que começam com a letra S. O código abaixo nos permite comparar a distribuição espacial de estabelecimentos de saúde de baixa complexidade (S002) e alta complexidade (S004).

```

sau_b <- ggplot(df) +
  geom_sf(aes(fill = as.factor(S002)), color = NA, alpha = 0.9) +
  scale_fill_brewer(palette = "YlGnBu", direction = 1, limits = factor(0:4)) +
  labs(
    title = "Estabelecimentos de saúde",
    subtitle = "Baixa complexidade",
    fill = "N. de estabelecimentos"
  ) +

```

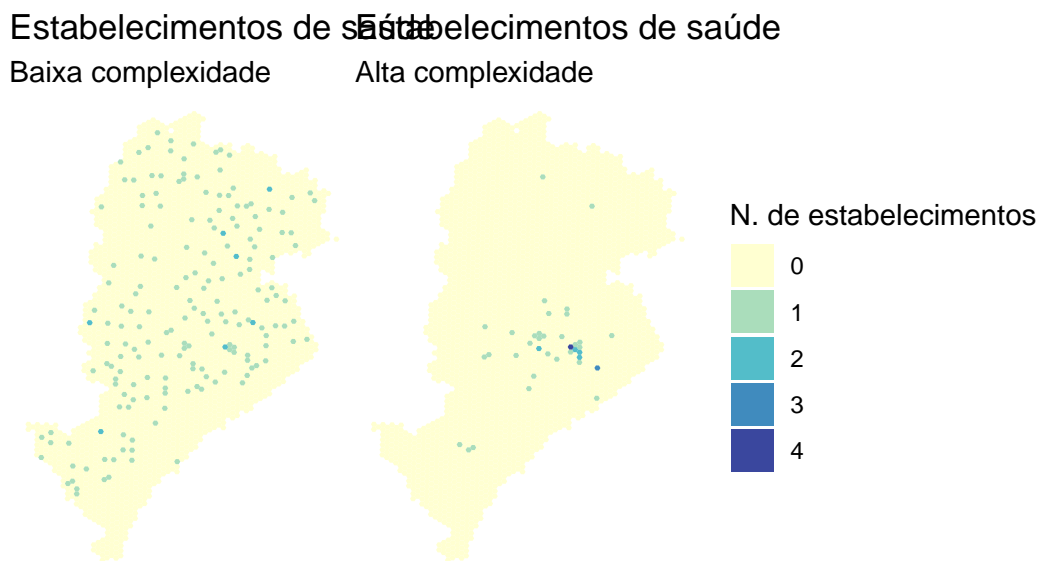
```

theme_void()

sau_a <- ggplot(df) +
  geom_sf(aes(fill = as.factor(S004)), color = NA, alpha = 0.9) +
  scale_fill_brewer(palette = "YlGnBu", direction = 1, limits = factor(0:4)) +
  labs(
    title = "Estabelecimentos de saúde",
    subtitle = "Alta complexidade",
    fill = "N. de estabelecimentos"
  ) +
  theme_void()

sau_b + sau_a + plot_layout(guides = 'collect')

```



8.5 Mapa de CRAS

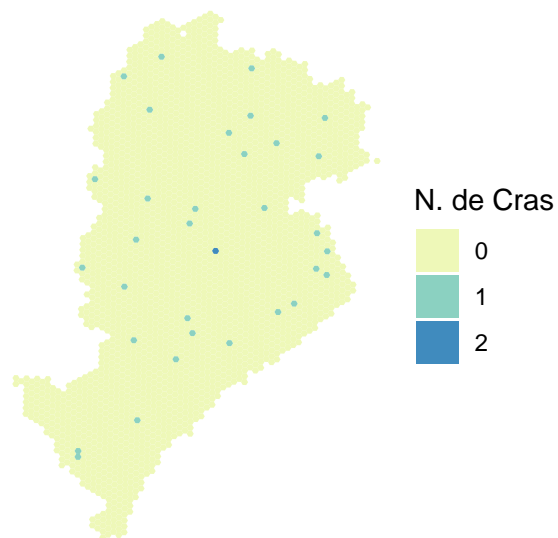
Por fim, a distribuição espacial dos Centros de Referência de Assistência Social (CRAS) pode ser analisada com a variável C001.


```

ggplot(df) +
  geom_sf(aes(fill = as.factor(C001)), color = NA, alpha = 0.9) +
  scale_fill_brewer(palette = "YlGnBu", direction = 1) +
  labs(
    title = "Centros de Referência de Assistência Social (Cras)",
    subtitle = "Belo Horizonte",
    fill = "N. de Cras"
  ) +
  theme_void()

```

Centros de Referência de Assistência Social (Cras) Belo Horizonte



9 Estimativas e mapas de acessibilidade

9.1 Download dos dados

Finalmente, o pacote `{aopdata}` também permite baixar, para todas cidades incluídas no projeto, estimativas anuais de acesso a empregos, serviços saúde, educação e assistência social por modo de transporte

Todos esses dados podem ser baixados com a função `read_access()`, que funciona da mesma maneira que as funções `read_population()` e `read_landuse()` apresentadas nos capítulos anteriores. Aqui, no entanto, além de indicar a cidade (`city`) e o ano (`year`) de referência para baixar os dados, você também precisa informar qual o modo de transporte (`mode`) será baixado e se você quer estimativas de acessibilidade no horário de pico (`peak = TRUE`) ou fora-pico (`peak = FALSE`). Esses dados representam a acessibilidade mediana entre 6h e 8h e entre 14h e 16h, respectivamente.

Neste exemplo, abaixo, nós mostramos como baixar os dados de acessibilidade urbana no ano de 2019 para São Paulo no período de pico. Nesse exemplo, nós baixamos tanto as estimativas de acessibilidade por automóvel quanto por transporte público, e empilhamos os dados num único `data.frame`. Note que essa função também baixa os dados de população e de uso do solo automaticamente.

```
library(aopdata)

# download aop accessibility data
df_pt <- read_access(
  city = "São Paulo",
  mode = "public_transport",
  year = 2019,
  peak = TRUE,
  geometry = TRUE,
  showProgress = FALSE
)

df_car <- read_access(
  city = "São Paulo",
```

```

mode = "car",
year = 2019,
peak = TRUE,
geometry = TRUE,
showProgress = FALSE
)

# row bind into a single data.frame
df <- rbind(df_pt, df_car)
head(df)

```

Simple feature collection with 6 features and 205 fields

Geometry type: POLYGON

Dimension: XY

Bounding box: xmin: -46.63863 ymin: -23.71413 xmax: -46.62834 ymax: -23.70485

Geodetic CRS: WGS 84

	id_hex	abbrev_muni	name_muni	code_muni	year	P001	P002	P003	P004	P005
1	89a81000003ffff	spo	Sao Paulo	3550308	2019	322	127	190	0	5
2	89a81000007ffff	spo	Sao Paulo	3550308	2019	16	3	13	0	0
3	89a8100000bffff	spo	Sao Paulo	3550308	2019	2386	1142	1232	2	10
4	89a8100000fffff	spo	Sao Paulo	3550308	2019	885	260	622	0	3
5	89a81000013ffff	spo	Sao Paulo	3550308	2019	725	340	380	0	5
6	89a81000017ffff	spo	Sao Paulo	3550308	2019	211	110	98	0	3

	P006	P007	P010	P011	P012	P013	P014	P015	P016	R001	R002	R003	T001	T002	T003
1	158	164	33	63	28	35	85	75	3	477.6	2	3	72	19	50
2	9	7	2	2	2	3	2	5	0	65.7	1	1	0	0	0
3	1216	1170	247	410	174	269	682	572	32	377.2	1	1	52	6	24
4	460	425	96	124	88	127	175	266	9	363.9	1	1	0	0	0
5	371	354	85	129	51	78	196	173	13	503.5	2	3	0	0	0
6	104	107	20	33	16	25	52	59	6	687.9	3	6	113	18	65

	T004	E001	E002	E003	E004	M001	M002	M003	M004	S001	S002	S003	S004	C001
1	3	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	22	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	30	2	0	2	1	1477	0	1168	309	0	0	0	0	0

	mode	peak	CMATT15	CMATB15	CMATM15	CMATA15	CMAST15	CMASB15	CMASM15
1	public_transport	1	315	71	175	69	0	0	0
2	public_transport	1	313	70	174	69	0	0	0
3	public_transport	1	229	43	161	25	2	2	2
4	public_transport	1	551	123	343	85	3	3	3

5	public_transport	1	421	90	262	69	2	2	2
6	public_transport	1	312	76	187	49	0	0	0
	CMASA15	CMAET15	CMAEI15	CMAEF15	CMAEM15	CMAMT15	CMAMI15	CMAMF15	CMAMM15
1	0	2	0	2	1	1477	0	1168	309
2	0	2	0	2	1	1477	0	1168	309
3	0	0	0	0	0	0	0	0	0
4	0	4	1	3	2	3132	460	1965	707
5	0	2	0	2	1	1477	0	1168	309
6	0	2	0	2	1	1477	0	1168	309
	CMACT15	CMPPPT15	CMPPPH15	CMPPPM15	CMPPPB15	CMPPPA15	CMPPPI15	CMPPPN15	CMPP0005I15
1	0	4895	2428	2467	2238	27	2	2628	505
2	0	4895	2428	2467	2238	27	2	2628	505
3	0	9537	4645	4892	4685	64	3	4785	872
4	0	13109	6331	6778	6453	61	8	6587	1088
5	0	4934	2452	2482	2368	48	0	2518	461
6	0	3199	1608	1591	1567	20	0	1612	317
	CMPP0614I15	CMPP1518I15	CMPP1924I15	CMPP2539I15	CMPP4069I15	CMPP70I15	CMATT30		
1	809	371	566	1313	1247	84	4797		
2	809	371	566	1313	1247	84	4228		
3	1538	726	1138	2550	2539	174	2648		
4	2012	967	1575	3383	3783	301	6200		
5	749	373	587	1271	1372	121	5773		
6	487	239	374	807	891	84	6608		
	CMATB30	CMATM30	CMATA30	CMAST30	CMASB30	CMASM30	CMASA30	CMAET30	CMAEI30
1	1293	2789	715	5	5	5	0	18	5
2	1170	2452	606	4	4	4	0	13	4
3	714	1526	408	4	4	4	0	9	2
4	1598	3641	961	5	5	5	0	19	6
5	1589	3271	913	5	5	5	0	18	5
6	1821	3659	1128	6	6	6	0	22	5
	CMAEF30	CMAEM30	CMAMT30	CMAMI30	CMAMF30	CMAMM30	CMACT30	CMPPPT30	CMPPPH30
1	13	4	15148	2409	10765	1974	1	59041	28457
2	9	3	10497	1949	6931	1617	1	40049	19392
3	7	3	7222	1069	5137	1016	1	59792	28974
4	13	5	15710	2527	10551	2632	2	101380	48709
5	13	4	15148	2409	10765	1974	1	81985	39473
6	17	5	19547	2409	14343	2795	1	83186	40111
	CMPPPM30	CMPPPB30	CMPPPA30	CMPPPI30	CMPPPN30	CMPP0005I30	CMPP0614I30	CMPP1518I30	
1	30584	28535	446	29	30031	5263	8809	4111	
2	20657	19159	273	14	20603	3610	6061	2866	
3	30818	28424	365	25	30978	5376	9204	4241	
4	52671	47601	955	66	52758	9076	15239	6971	
5	42512	39784	695	49	41457	7189	12278	5675	

6	43075	40359	728	51	42048	7302	12371	5749	
	CMPP1924I30	CMPP2539I30	CMPP4069I30	CMPP70I30	CMATT60	CMATB60	CMATM60	CMATA60	
1	6817	15664	16733	1644	96195	18711	54773	22711	
2	4673	10486	11274	1079	72810	14161	40949	17700	
3	6927	15859	16646	1539	147237	26399	82208	38630	
4	11658	27259	28303	2874	107938	21292	61768	24878	
5	9311	21712	23351	2469	161353	28971	91687	40695	
6	9459	22070	23710	2525	158016	28997	91280	37739	
	CMAST60	CMASB60	CMASM60	CMASA60	CMAET60	CMAEI60	CMAEF60	CMAEM60	CMAMT60
1	50	38	48	6	120	40	76	29	87761
2	40	34	39	2	108	36	68	28	79902
3	57	45	56	5	135	45	88	30	93958
4	56	43	54	6	128	42	82	31	92799
5	59	44	57	7	141	47	90	32	97925
6	56	43	54	6	139	47	87	32	96281
	CMAMI60	CMAMF60	CMAMM60	CMACT60	CMPPT60	CMPPH60	CMPPM60	CMPPB60	CMPPA60
1	15059	56558	16144	5	547471	259971	287500	277734	11134
2	14014	50455	15433	5	501558	238550	263008	247962	8781
3	16318	61918	15722	6	633994	299454	334540	339064	19838
4	15615	60049	17135	5	689942	325560	364382	378571	21810
5	16744	63838	17343	6	666492	314759	351733	360532	20661
6	16812	62265	17204	6	632529	299225	333304	334655	17957
	CMPP160	CMPPN60	CMPP0005I60	CMPP0614I60	CMPP1518I60	CMPP1924I60	CMPP2539I60		
1	407	258196	44866	76195	34899	59393	146870		
2	383	244432	41903	71199	32501	54873	134759		
3	473	274619	49586	83666	38585	67298	169669		
4	496	289065	53223	89680	41446	72354	184045		
5	483	284816	51720	87430	40333	70418	178015		
6	451	279466	49804	84545	38888	67458	168928		
	CMPP4069I60	CMPP70I60	CMATT90	CMATB90	CMATM90	CMATA90	CMAST90	CMASB90	CMASM90
1	162048	23200	1168463	166686	577124	424653	169	111	160
2	146361	19962	1088120	153982	540566	393572	160	107	152
3	194234	30956	1428750	211239	719220	498291	190	119	179
4	214049	35145	1362473	191461	669950	501062	182	117	172
5	205504	33072	1443193	211449	723001	508743	203	131	191
6	192822	30084	1406215	203383	701014	501818	204	132	193
	CMASA90	CMAET90	CMAEI90	CMAEF90	CMAEM90	CMAMT90	CMAMI90	CMAMF90	CMAMM90
1	42	337	107	207	92	218542	30864	131526	56152
2	39	328	105	201	90	213963	30365	128917	54681
3	50	370	123	222	101	232126	34693	138061	59372
4	46	355	117	214	92	223492	32779	134438	56275
5	51	415	136	248	114	264729	38719	158252	67758
6	51	415	137	247	114	266946	39578	159137	68231

	CMACT90	CMPTT90	CMPPH90	CMPPM90	CMPPB90	CMPPA90	CMPP190	CMPPN90	CMPP0005I90
1	10	1697502	790958	906544	1113577	80011	1336	502578	111963
2	9	1617428	753988	863440	1054271	74800	1240	487117	107419
3	10	1918445	894649	1023796	1274988	90290	1602	551565	125137
4	10	2096063	978494	1117569	1384576	94542	2065	614880	138110
5	12	2027043	946208	1080835	1333715	91633	1714	599981	133501
6	12	1950319	910888	1039431	1272271	88334	1558	588156	129637
	CMPP0614I90	CMPP1518I90	CMPP1924I90	CMPP2539I90	CMPP4069I90	CMPP70I90			
1	183412	87698	165449	454122	573567	121291			
2	176240	84154	157827	432825	545306	113657			
3	204140	97829	187004	513061	650643	140631			
4	225800	107835	205069	561408	707435	150406			
5	218843	104726	198411	541774	684520	145268			
6	212862	101535	191408	521731	655761	137385			
	CMATT120	CMATB120	CMATM120	CMATA120	CMAST120	CMASB120	CMASM120	CMASA120	
1	2136716	328724	1094450	713542	357	235	338	82	
2	2095040	321900	1068319	704821	355	235	337	80	
3	2203072	344047	1136019	723006	374	243	356	86	
4	2494268	397258	1308021	788989	418	266	395	97	
5	2275396	357833	1177979	739584	401	265	381	90	
6	2195795	342423	1130056	723316	387	257	367	88	
	CMAET120	CMAEI120	CMAEF120	CMAEM120	CMAMT120	CMAMI120	CMAMF120	CMAMM120	
1	821	288	481	221	518042	83000	313683	121359	
2	815	288	479	216	516846	83610	313657	119579	
3	855	293	508	230	532772	82410	323147	127215	
4	976	340	576	260	610890	97742	369803	143345	
5	948	333	559	251	600516	97066	367244	136206	
6	942	334	554	254	604830	98765	368433	137632	
	CMACT120	CMPTT120	CMPPH120	CMPPM120	CMPPB120	CMPPA120	CMPP1120	CMPPN120	
1	20	4018897	1890513	2128384	2540395	134426	4039	1340037	
2	20	3860887	1814742	2046145	2445536	130803	3868	1280680	
3	21	4320562	2032079	2288483	2745882	142238	4320	1428122	
4	24	4752649	2234731	2517918	3037441	153875	5098	1556235	
5	22	4351102	2049261	2301841	2735091	140022	4377	1471612	
6	21	4191716	1974700	2217016	2627575	136196	4172	1423773	
	CMPP0005I120	CMPP0614I120	CMPP1518I120	CMPP1924I120	CMPP2539I120	CMPP4069I120			
1	281477	467671	217602	408477	1084792	1300130			
2	269676	448086	208699	391944	1041317	1251187			
3	301337	501274	233530	439651	1163931	1400822			
4	330744	549438	256378	484187	1277780	1541189			
5	307070	511736	237911	444278	1173140	1401287			
6	297029	494835	229726	427661	1129908	1348228			
	CMPP70I120	TMIST	TMISB	TMISM	TMISA	TMJET	TMIEI	TMIEF	TMIEI

1	258748	17	17	17	53	8.0	26	8.0	8.0	21
2	249978	21	21	21	55	9.0	28	9.0	9.0	23
3	280017	13	13	13	50	17.0	26	17.0	17.0	30
4	312933	13	13	13	53	7.0	13	7.0	7.0	28
5	275680	12	12	12	51	8.0	24	8.0	8.0	19
6	264329	16	16	16	49	5.8	22	5.8	5.8	16

geometry

```
1 POLYGON ((-46.63251 -23.711...
2 POLYGON ((-46.63552 -23.709...
3 POLYGON ((-46.62941 -23.709...
4 POLYGON ((-46.63242 -23.708...
5 POLYGON ((-46.6326 -23.7141...
6 POLYGON ((-46.63561 -23.712...
```

Os nomes das variáveis (colunas) com estimativas de acessibilidade também estão organizadas com códigos, como CMAEF30, TMISB ou CMPPM60. O nome das colunas com estimativas de acessibilidade são a junção de três componentes: o tipo de indicador de acessibilidade; o tipo de oportunidade / pessoas; e o tempo limite.

1. O **tipo de indicador** de acessibilidade é indicado pelas primeiras 3 letras. O projeto AOP atualmente inclui três tipos de indicadores:

- CMA - Indicador de acessibilidade cumulativo ativo
- CMP - Indicador de acessibilidade cumulativo passivo
- TMI - Indicador de tempo mínimo até oportunidade mais próxima

2. O **tipo de atividade** é indicado pelas letras seguintes, no meio do nome da variável. O projeto AOP atualmente inclui diversos tipos de atividades:

- TT - Todos empregos
- TB - Empregos de baixa escolaridade
- TM - Empregos de média escolaridade
- TA - Empregos de alta escolaridade
- ST - Todos estabelecimentos de saúde
- SB - Estabelecimentos de saúde de baixa complexidade
- SM - Estabelecimentos de saúde de média complexidade
- SA - Estabelecimentos de saúde de alta complexidade
- ... e assim por diante.

No caso do indicador de acessibilidade passiva (CMP), as letras do meio do nome da variável indicam qual o grupo populacional de referência.

- PT - População total
- PH - População de homens

- PM - População de mulheres
- PB - População branca
- PN - População negra
- P1924I - População de 19 a 24 anos de idade
- P2539I - População de 25 a 39 anos de idade

3. O **tempo limite de viagem** é indicado pelos números no final do nome da variável. Esses números somente se aplicam para os indicadores de acessibilidade cumulativa ativa (CMA) e passiva (CMP).

Exemplos:

CMAEF30: Número de escolas de ensino fundamental acessíveis em até 30 minutos

TMISB: Tempo de viagem até o estabelecimento de saúde mais próximo com serviços de baixa complexidade

CMPPM60: Quantidade de mulheres que conseguem acessar determinado hexágono em até 60 minutos

Lembre-se, a descrição completa do dicionário de variáveis está disponível no [site to pacote aopdata](#).

A seguir, nós mostramos alguns exemplos de como visualizar essas estimativas de acessibilidade.

9.2 Mapa do tempo para acessar o hospital mais próximo

Neste exemplo, nós vamos comparar o nível de acessibilidade até hospitais entre modos automóvel *vs* transporte público. Para analisar qual o tempo mínimo de viagem (TMI) até um hospital de alta complexidade (SA), nós precisamos analisar a variável **TMISA**. Com o código abaixo, nós carregamos as bibliotecas para manipulação e visualização de dados, e visualizamos a distribuição espacial dos valores de **TMISA** para ambos modos de transporte.

Note, no entanto, que os tempos de viagem por transporte público costumam ser muito mais longos do que por automóvel. Então para facilitar a visualização dos resultados, nós truncamos a distribuição dos valores de **TMISA** em 60 minutos ou mais.

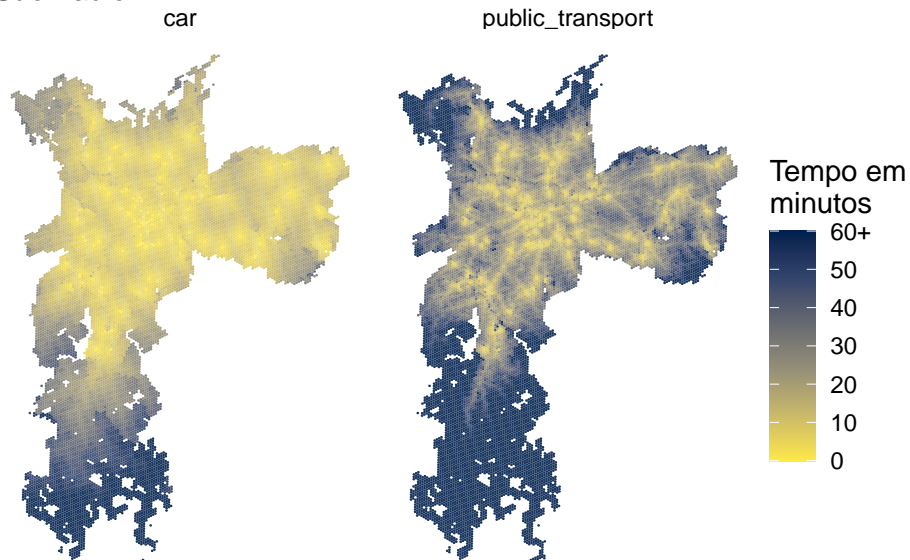
```
library(ggplot2)
library(data.table)
library(patchwork)
library(scales)
library(sf)
```



```
# truncate max values to 60 min.
df$TMISA <- ifelse(df$TMISA > 60, 60, df$TMISA)

ggplot(subset(df, !is.na(mode))) +
  geom_sf(aes(fill = TMISA), color = NA, alpha = 0.9) +
  scale_fill_viridis_c(
    option = "cividis",
    direction = -1,
    breaks = seq(0, 60, 10),
    labels = c(seq(0, 50, 10), "60+")
  ) +
  labs(
    title = "Tempo de viagem até hospital de alta complex. mais próximo",
    subtitle = "São Paulo",
    fill = "Tempo em\nminutos"
  ) +
  facet_grid(. ~ mode) +
  theme_void()
```

Tempo de viagem até hospital de alta complex. mais próximo
São Paulo



9.3 Mapa da quantidade de oportunidades acessíveis

Uma vez que os dados já foram baixados do {aopdata}, é muito simples comparar por exemplo a quantidade de oportunidades acessíveis em diferentes tempos de viagem. O código abaixo ilustra como visualizar num mapa o número de empregos acessíveis em até 60 e 90 minutos de viagem por transporte público, e como colocar esses mapas lado a lado para comparação.

```
# limits for legend scale
value_limits <- c(0, max(df_pt$CMATT90, na.rm=T) / 1000)

# create maps
fig60 <- ggplot(subset(df_pt, !is.na(mode))) +
  geom_sf(aes(fill = CMATT60 / 1000), color = NA, alpha = 0.9) +
  scale_fill_viridis_c(option = "inferno", limits = value_limits) +
  labs(subtitle = "em até 60 min.", fill = "Empregos") +
  theme_void()

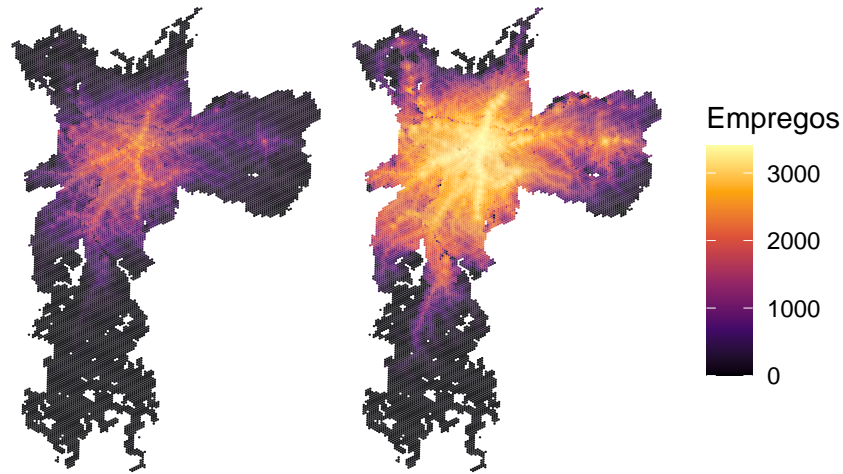
fig90 <- ggplot(subset(df_pt, !is.na(mode))) +
  geom_sf(aes(fill = CMATT90 / 1000), color = NA, alpha = 0.9) +
  scale_fill_viridis_c(option = "inferno", limits = value_limits) +
  labs(subtitle = "em até 90 min.", fill = "Empregos") +
  theme_void()

fig60 +
  fig90 +
  plot_layout(guides = "collect") +
  plot_annotation(
    title = "Quantidade de empregos acessíveis por transporte público",
    subtitle = "São Paulo"
  )
```

Quantidade de empregos acessíveis por transporte público São Paulo

em até 60 min.

em até 90 min.



9.4 Desigualdades de acesso a oportunidades

Existem diversas maneiras de se analisar quão desiguais são as condições de acesso a oportunidades a partir dos dados do {aopdata}. Nós apresentamos nesta subseção três exemplos de como esse tipo de análise pode ser realizado.

Desigualdade no tempo de acesso TMI

Neste primeiro exemplo, nós vamos comparar qual o tempo médio de viagem até o hospital mais próximo para pessoas de diferentes níveis de renda. Para isso, o código abaixo calcula o valor médio de **TMISA** ponderada pela população em cada hexágono. Essa ponderação é necessária porque o número de hexágonos pode variar muito entre hexágonos.

Antes disso, cabe observar que alguns hexágonos da cidade não conseguem acessar nenhum hospital em até 2h de viagem. Em casos como esse, o valor das variáveis **TMI__** é infinito (**Inf**). Para lidar com esses casos, nós substituímos abaixo todos valores **Inf** por 120 minutos.

```
# copy data to new data.table
dt <- copy(df_pt)
setDT(dt)

# replace Inf travel time with 120
```

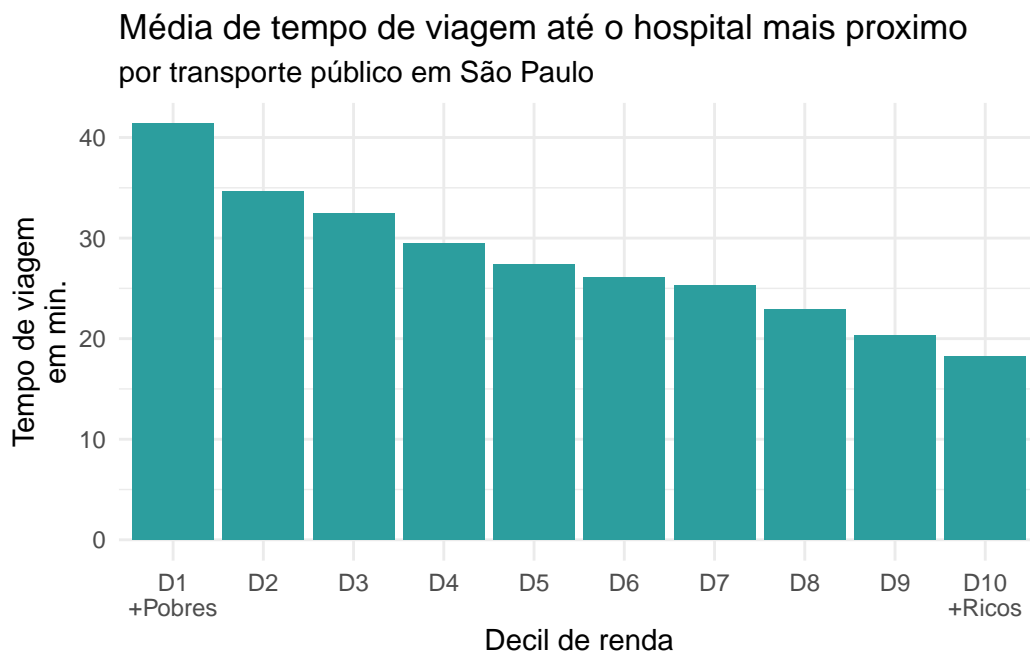
```

dt[, TMISA := fifelse(TMISA==Inf, 120, TMISA)]

# calculate average travel time by income
temp <- dt[
  ,
  .(average = weighted.mean(x = TMISA, w = P001, na.rm = TRUE)),
  by = R003
]
temp <- na.omit(temp)

ggplot(temp) +
  geom_col(aes(y = average, x = factor(R003)), fill = "#2c9e9e", color = NA) +
  scale_x_discrete(labels = c("D1\n+Pobres", paste0("D", 2:9), "D10\n+Ricos")) +
  labs(
    title = "Média de tempo de viagem até o hospital mais proximo",
    subtitle = "por transporte público em São Paulo",
    x = "Decil de renda",
    y = "Tempo de viagem\nem min."
  ) +
  theme_minimal()

```

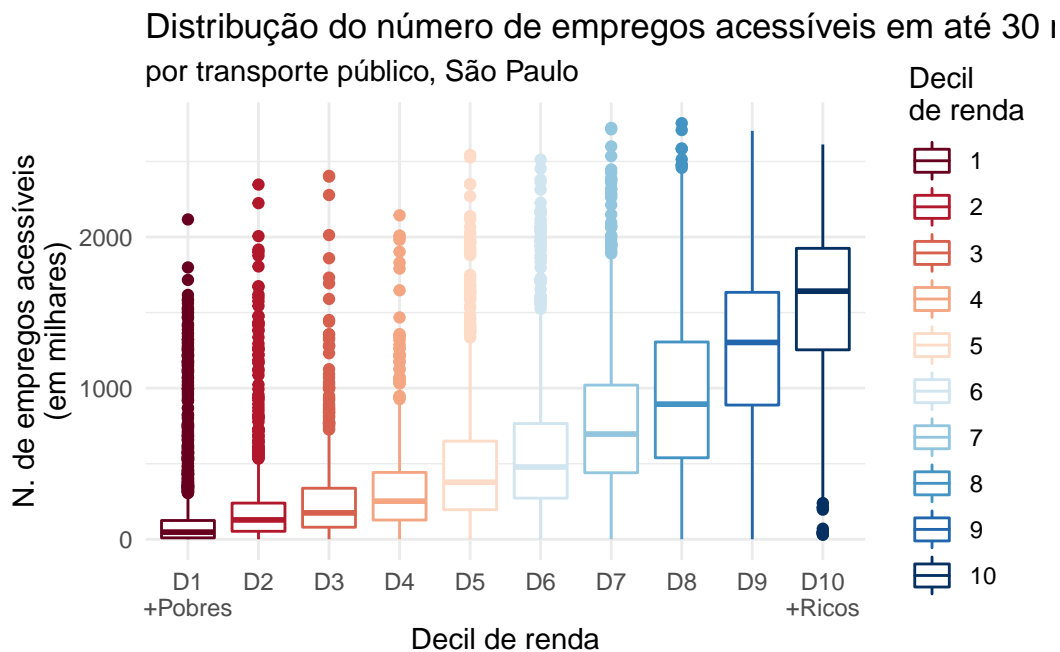


Desigualdade do número de oportunidades acessíveis CMA

Outra maneira de examinar a desigualdade de acesso a oportunidades é comparar a quantidade de oportunidades acessíveis por diferentes grupos populacionais considerando-se um mesmo modo de transporte e tempo de viagem. Nesse caso, nós analisamos o Indicador de acessibilidade cumulativo ativo (CMA).

Neste exemplo abaixo, nós utilizamos *box plots* para comparar a quantidade de empregos acessíveis por transporte público em até 30 minutos de viagem.

```
ggplot(subset(dt, !is.na(R003))) +
  geom_boxplot(
    aes(x = factor(R003), y = CMATT60 / 1000, color = factor(R003))
  ) +
  scale_color_brewer(palette = "RdBu") +
  labs(
    title = "Distribuição do número de empregos acessíveis em até 30 min.",
    color = "Decil\nde renda",
    subtitle = "por transporte público, São Paulo",
    x = "Decil de renda",
    y = "N. de empregos acessíveis\n(em milhares)"
  ) +
  scale_x_discrete(labels = c("D1\n+Pobres", paste0("D", 2:9), "D10\n+Ricos")) +
  theme_minimal()
```



Também é possível comparar como diferentes modos de transporte possibilitam diferentes níveis de acessibilidade, e como essa diferença pode variar muito entre cidades. Nesse exemplo, abaixo, nós vamos comparar a quantidade de empregos acessíveis em até 30 minutos de viagem a pé e de carro. O primeiro passo é baixar os dados de acessibilidade a pé e por automóvel para todas as cidades.

```
df_car <- read_access(  
  city = "all",  
  mode = "car",  
  year = 2019,  
  showProgress = FALSE  
)  
  
df_walk <- read_access(  
  city = "all",  
  mode = "walk",  
  year = 2019,  
  showProgress = FALSE  
)
```

O próximo passo é calcular para cada cidade a média ponderada do número de empregos acessíveis em até 30 minutos (CMATT30) para cada um dos modos de transporte. Com esses resultados prontos para cada modo, nós juntamos essas estimativas num único `data.frame` e calculamos a razão da acessibilidade por carro dividida pela acessibilidade à pé.

```
# calculate the average number of jobs accessible in 30min  
df_car2 <- df_car[  
  ,  
  .(access_car = weighted.mean(CMATT30, w = P001, na.rm = TRUE)),  
  by = name_muni  
]  
  
df_walk2 <- df_walk[  
  ,  
  .(access_walk = weighted.mean(CMATT30, w = P001, na.rm = TRUE)),  
  by = name_muni  
]  
  
# merge and get the ratio between walking and driving access  
df <- merge(df_car2, df_walk2)  
df[, ratio := access_car / access_walk]
```

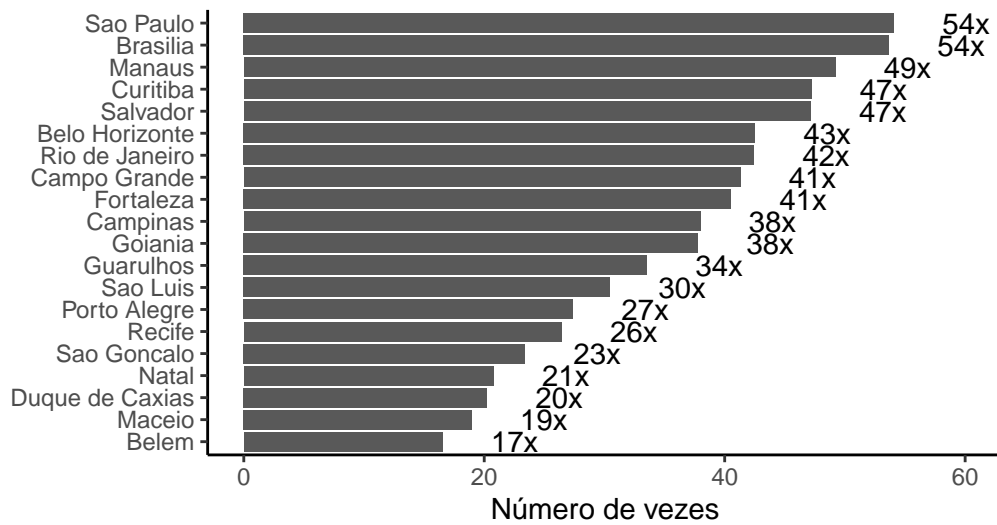
```
head(df)
```

	name_muni	access_car	access_walk	ratio
1:	Belem	155270.4	9392.235	16.53179
2:	Belo Horizonte	529890.0	12464.233	42.51284
3:	Brasilia	220575.9	4110.703	53.65892
4:	Campinas	256333.1	6748.923	37.98133
5:	Campo Grande	172680.5	4181.209	41.29919
6:	Curitiba	494376.9	10471.135	47.21331

Pronto, agora é só visualizar os resultados. Como esperado, a figura abaixo mostra que é possível acessar muito mais empregos em 30 min. por automóvel do que em caminhada. No entanto, essa diferença varia muito entre cidades. Em São Paulo e em Brasília, 30 minutos de viagem por automóvel permite acessar, em média, um número de empregos 54 vezes maior do que a pé. Em Belém, onde observamos a menor diferença, o automóvel ainda permite acessar 17 vezes mais empregos do que a caminhada.

```
ggplot(df, aes(x = ratio, y = reorder(name_muni, ratio))) +  
  geom_bar(stat = "identity") +  
  geom_text(aes(x = ratio + 6 , label = paste0(round(ratio), "x"))) +  
  labs(  
    title = "Diferença da quantidade de empregos acessíveis por\nautomóvel vs a pé",  
    subtitle = "em até 30 min. de viagem",  
    y = "",  
    x = "Número de vezes"  
  ) +  
  theme_classic()
```

Diferença da quantidade de empregos acessíveis por
automóvel vs a pé
em até 30 min. de viagem



Referências bibliográficas

- Luo, Wei, and Fahui Wang. 2003. “Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region.” *Environment and Planning B: Planning and Design* 30 (6): 865–84. <https://doi.org/10.1068/b29120>.
- Paez, Antonio, Christopher D. Higgins, and Salvatore F. Vivona. 2019. “Demand and Level of Service Inflation in Floating Catchment Area (FCA) Methods.” Edited by Tayyab Ikram Shah. *PLOS ONE* 14 (6): e0218773. <https://doi.org/10.1371/journal.pone.0218773>.
- Pereira, R. H. M., Carlos Kaue Vieira Braga, Bernardo Serra, and Vanessa Nadalin. 2019. “Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras, 2019.” {Texto para Discussão} 2535. Instituto de Pesquisa Econômica Aplicada (Ipea).
- Pereira, Rafael H. M., Daniel Herszenhut, Carlos Kauê Vieira Braga, João Pedro Bazzo, João Lucas Albuquerque Oliveira, João Pedro Parga, Marcus Saraiva, Luiz Pedro Silva, Lucas Warwar, and Diego Bogado Tomasiello. 2022. “Distribuição espacial de características sociodemográficas e localização de empregos e serviços públicos das vinte maiores cidades do Brasil.” *Texto para Discussão IPEA* 2772 (June): 28. <https://doi.org/10.38116/td2772>.